ENCYCLOPEDIA OF Biodiversity

VOLUME 1

ACADEMIC PRESS

Editor-in-Chief

Simon Asher Levin, *Moffett Professor of Biology,*
Princeton University, Princeton, New Jersey, USA

Associate Editors

International Editorial Advisors

Foreword

Biodiversity is the totality of the inherited variation of all forms of life across all levels of variation, from ecosystem to species to gene. Soon after the term was introduced at the first National Forum on Biodiversity in 1986, and after it began its rapid spread around the world, there occurred a reconfiguration in the way much of the science of biology is conceived. Where previously comparative biology had been almost entirely focused on the fundamentals of classification, evolution, behavior, and ecology, now it was augmented by a wide range of analyses from the social sciences. Where taxonomy and biogeography had been marginalized through the middle half of the twentieth century, now they moved back toward center stage. And where extinction had been little more than a phenomenon recognized and lamented, now it became a major concern of science. Much of ecology shifted toward the study of biodiversity's role in the assembly and maintenance of ecosystems. A growing number of economists, political scientists, and bioethicists took up the issue as part of their scholarly agenda. From this mix the discipline of conservation biology was born, and the Society of Conservation Biology became one of the fastest growing organizations in modern science. In 1992 the Rio Summit catapulted biodiversity to global prominence, from which most of the nations of the world endorsed the Convention on Biodiversity and have since used it as a guideline for conservation programs.

The new biodiversity initiative gave organismic and evolutionary biologists a global mission worthy of their science. It confirmed for those who labored in the vineyards that, as medicine is to molecular and cellular biology, the environment is to organismic and evolutionary biology. The first is responsible for personal health, and the second for planetary health. The additional evidence adduced moreover made clear that we are in the midst of an episode of massive extinction, unprecedented since that closing of the Mesozoic Era—and that scientists must lead the attempt to save the Creation.

The articles in the *Encyclopedia of Biodiversity* are unusually eclectic, yet organized by a set of easily articulated goals. They are the following: to carry the systematics and biogeography of the world fauna and flora toward completion; map the hot spots where conservation will save the most biodiversity; orient studies of natural history to understand and save threatened species; advance ecosystems studies and biogeography to create the needed principles of community assembly and maintenance; acquire the knowledge of resource use, economics, and polity to advance conservation programs based on sustainability; and enrich the ethic of global conservation in terms persuasive to all.

The road ahead, down which we must urgently travel, will be smoothed by the exponential growth of information and a growing public awareness and support. The *Encyclopedia of Biodiversity* will serve as an important knowledge base to guide this supremely important effort.

Edward O. Wilson
Museum of Comparative Zoology
Harvard University

Preface

The science of biodiversity has become the science of our future. Our awareness of the disappearance of biodiversity has brought with it a long-overdue appreciation of the magnitude of our loss, and a determination to develop the tools to protect our future. This encyclopedia brings together, for the first time in its completeness, study of the dimensions of diversity with examination of the services that biodiversity provides, and measures to protect it.

The entries in the encyclopedia have been arranged alphabetically, but the coverage is designed functionally. At the core is a comprehensive survey of biodiversity, across taxonomic groups and ecological regions. The emergence of biodiversity is then placed in an evolutionary perspective, as background for an understanding of current trends. Particular attention is given to the loss of services—for example, in fisheries, forestry and climate mediation—that are derived from natural systems. These are placed in an economic framework through a comprehensive set of papers that address problems of valuation, costs, and benefits, and develop a framework for prioritizing actions. Finally, a review is given of institutions and other mechanisms that exist and are needed for the preservation of biodiversity and, with it, the services that humans derive from nature.

The background for understanding biodiversity is to be found in the fossil record, and in the evolutionary patterns and trends that it reveals. The encyclopedia hence discusses these patterns, the origins of biodiversity, the effects of geological events, the mechanisms of evolution, and the uniqueness of the evolutionary process, with implications for conservation and restoration. The essential processes in macroevolution are those of speciation and extinction, which together govern the dynamics of diversity at higher levels of organization. These are given extensive coverage, both from a mechanistic and from a historical perspective, and provide an essential context for understanding the rest of the contributions.

The classification of organisms into species and higher taxa, and the elucidation of the mechanisms of natural selection, were the essential intellectual advances that allowed the development of the science of biodiversity. Carl Linnaeus introduced a systematic framework for understanding phylogenies, which continues to provide the foundation for evolutionary studies today; and Charles Darwin's great legacy—the theory of evolution by natural selection—is the essential organizing principle for understanding the processes that gave rise to the patterns Linnaeus recognized. The encyclopedia provides unmatched taxonomic coverage of the organization of diversity into taxonomic groups and complements that with an extensive examination of ecosystems by biogeographic region and by functional type. These chapters elucidate latitudinal trends, life zones, species–area relationships and the distribution of diversity within and among ecological communities.

Throughout the core chapters, there is a healthy balance between empirical facts and conceptual theories. Such theories help to illuminate principles that cross systems and levels of organization, and transform the study of biodiversity into a science. Basic ecological constructs, such as the habitat and the niche, are given extensive treatment, as are key ecological mechanisms such as competition, predation, herbivory, parasitism and mutualism. These treatments are complemented by exploration of fundamental evolutionary mechanisms related to local differentiation, aspect diversity, sex, and recombination, and especially theories of extinction.

With these foundational chapters in hand, one can turn to the contemporary problems in biodiversity and compare today's rapid rates of change to the historical patterns. Key chapters examine agriculture, fisheries, and forests, their importance to human needs, and their status and trends in response to changing land-use patterns, population growth, overexploitation, and climate change. Threatened and endangered species are discussed in detail, with relation to the consequences of the spread of invading species.

The utilization of nature's bounty for food, fiber and fuel provides some of the most obvious benefits of biodiversity to humanity. Equally important, however, are the things that are less well appreciated: the potential for the discovery of new pharmaceuticals that can improve human health, the role of biodiversity in pollinating crops and wild species, and the importance of natural systems in regulating climate, mediating nutrient fluxes, and sequestering carbon as well as toxic materials. Each of these services provides humans with direct and indirect benefits, and somehow we must find ways to weigh these benefits, along with the ethical and aesthetic values we place on natural systems and biodiversity, to provide priorities for action. Only recently have economists recognized the importance of such issues as intellectual challenges essential to our survival on the planet. Much of biodiversity is exploited by humanity as part of a global commons, in which one does not pay in fair measure for extracting parts, or affecting the commons otherwise through land use or pollution. Economists have come to realize, along with ecologists and others, the magnitude of the externalities involved. When such externalities are involved, the market does not function as it must to maintain the resource, and new measures are needed if the sustainability and resilience of these resources are to be preserved. There is as yet, however, no ecological equivalent to the power of financial institutions, such as the Federal Reserve Board in the United States, to modify individual incentives sufficiently to maintain regional or global stability in the system of interest. In this encyclopedia, some of the most enlightened and thoughtful economists turn their attention to the economic challenges, and discuss the mechanisms and institutions that might be needed.

Together, the state-of-the-art entries in this encyclopedia tell an exciting story of how biodiversity arose, continues to arise, and is maintained. It is a story of a complex, self-organizing system—the biosphere—whose pieces can be examined individually, but cannot be understood outside the context of the whole. It is also a story of the coevolution of the biosphere and *Homo sapiens*, the first species whose own activities can feed back to influence the evolution of the biosphere on time scales that could lead to its own demise. The articles in the encyclopedia can be used as material for a wide spectrum of courses, tracing the history of the emergence of biodiversity from its origins to the challenges we face today.

This has been a massive effort, but one of the most rewarding I have ever undertaken. So many people have played a role that it is difficult to know where to begin. The project began through the initiative of Scott Bentley at Academic Press, and then was managed flawlessly by Chris Morris at AP. I cannot recall ever having dealt with an editor who operated more professionally than Chris, who combined a true vision and enthusiasm for the project with a sense of economic realities and the energy and insight to make the whole project work. In this he was ably assisted by outstanding Academic Press staff, especially Naomi Henning, Nick Panissidi, and Ann Marie Martin. At every step, it has been a pleasure to work with Academic Press, and I especially single out Chris for his fantastic and scholarly efforts.

At the next stage, the Editorial Boards were terrific in generating and commenting on ideas, suggesting authors, and critiquing contributions. More than 400 authors then adopted our view of the importance of the project, accepted the task of writing, and produced timely and comprehensive articles that make this Encyclopedia like no other source available today. To all of these, I extend my thanks and congratulations.

And finally, special acknowledgment and gratitude are due my wife, Carole, and my assistant, Amy Bordvik. Carole put up with the late nights and obsessiveness that were essential to the process, and Amy worked tirelessly and without complaint through the whole long process. To them, any expression of thanks is insufficient.

Simon Levin
Princeton University

References

Simon A. Levin (1999). *Fragile Dominion*, Perseus Books. Reading, Massachusetts

Edward O. Wilson (1992). *The Diversity of Life*. Norton. New York, New York.

Dedication

Dedicated to the memory of three encyclopedia authors, Takuya Abe, Masahiko Higashi, and Gary Polis, and their colleagues Shigeru Nakano and Michael Rose, who perished March 27, 2000 in a tragic boating accident while on a research trip in Baja California. Masahiko Higashi was also a member of the Board of International Editorial Advisors.

ACID RAIN AND DEPOSITION

George R. Hendrey
Brookhaven National Laboratory

GLOSSARY

acid deposition The combination of acid rain plus dry deposition; a term preferred over "acid rain."

acid rain Rain, fog, snow, sleet, or hail with pH less than 5.6.

aerosols Fine particulate matter suspended in the atmosphere, with diameters less than 5.5 μm.

alkalinity The acid-neutralizing capacity (ANC) of water: $ANC = [HCO_3^- + CO_3^{2-} + OH^-] - [H^+]$.

cation exchange capacity The total of exchangeable cations that a soil can absorb.

dry deposition Deposition of dry pollutants from the atmosphere including gases and aerosols.

macrophytes Vascular plants, mosses, liverworts, and macro-algae.

metric ton 1000 kg.

periphyton Community of organisms dominated by algae growing on submerged surfaces.

phytoplankton Microscopic plants that live suspended in the water column.

I. ACID DEPOSITION

A. Acid Rain

In the mid-1970s the existence of highly acidic rain became widely known because it appeared to be reducing biodiversity through acidification of surface waters. This ecological problem was linked to emissions of compounds of sulfur and nitrogen from fuel combustion that are oxidized in the atmosphere to form sulfuric acid (H_2SO_4) and nitric acid (HNO_3) and related compounds that make precipitation very acidic, commonly referred to as "acid rain." Large, national-scale research projects have since found that over large areas of eastern North America and northern Europe, the deposition of these acids and related substances has led to extensive acidification of lakes and streams and the extinction of populations of fish from many surface waters. High-elevation forests are injured by acid deposition and buildings and monuments are corroded. Phenomena related to acid deposition reduce atmospheric visibility and impact human health. This knowledge has led to the regulation of air pollutants that is effective in reducing some of these problems. The most comprehensive source of information on this subject is the report series of the U.S. National Acid Deposition Assessment Program (NAPAP) published in 1990.

B. Dry Deposition

Dry deposition occurs when, in the absence of condensed water droplets, acid-forming substances in the atmosphere are deposited as gases and dry particles. Dry deposition may be in the form of a gas, such as SO_2, or in the form of a fine, dry aerosol particle such as ammonium sulfate [$(NH_4)_2SO_4$]. In landscapes receiving this deposition, runoff water from acid rain adds to the dry-deposited materials, making the combination more acidic than the falling rain alone.

C. Acid Deposition

Acid deposition, a term preferred over acid rain, is the combination of acid rain plus dry deposition. The most important chemical species of acid deposition are hydrogen ion (H^+), oxides of sulfur (SO_x) and nitrogen (NO_x), including the strong acid anions sulfate (SO_4^{2-}), nitrate (NO_3^-), and chloride (Cl^-), and ammonium (NH_4^+). These substances are dissolved in liquid water (rain or fog) or adsorbed onto frozen water (snow, sleet, or hail) so that the hydrogen ions (H^+) are dissociated from the acid anions.

Controls on the emissions of SO_2 already in place in both North America and Europe are reducing acid deposition. NO_x emissions and deposition, however, continue to increase. With these two opposing trends, there has been only a slight decrease in the acidity of "acid rain."

D. Acidity and the pH Scale

Pure water is a very weak acid ($H_2O \rightarrow H^+ + OH^-$), and the concentrations of H^+ and OH^- are equal. The amount of H^+ present in pure water under standard conditions (20°C, 1 atm pressure) is 1 ten-millionth of a gram of H^+ in a liter of water (0.0000001 *M*), or 10^{-7} moles per liter of water (mol/liter). Acidity is measured on the pH scale expressed as the negative logarithm of the H^+ concentration. Thus, pure water has a pH of 7. An acid concentration 10 times greater than pure water can occur if acid-forming anions are present. This solution will have one-millionth of a gram of H^+ in water, or 10^{-6} mol/liter, and the pH is 6. Thus, each whole pH unit lower represents a 10-fold increase in acidity. Over most of the eastern United States and other areas receiving acid deposition the pH of rain is in the range 4.1–4.8. Of the anions associated with precipitation acidity, SO_4^{2-} accounts for about 60% and NO_3^- for about 40%.

II. CAUSES OF ACID RAIN

Acid deposition has been occurring for a long time. In 1856, Robert Angus Smith, who was chief alkali inspector for Britain, wrote, "It has often been observed that the stones and bricks of buildings, especially under projecting parts, crumble more readily in large towns where coal is burnt. . . . I was led to attribute this effect to the slow but constant action of acid rain." Smith was concerned about air pollution and soot in Manchester, England. In the mid-nineteenth century, sulfurous fumes from the burning of coal in homes and factories reacted with water in the air to produce a dilute solution of sulfuric acid that attacked limestone and lime-based mortar in brickwork. Smith's acid rain problems tended to be local in scale. Chimneys in those days were low and their smoke spread out at low elevation across cities and towns. The problem that Smith described led to a gradual increase in the heights of smoke-stacks to allow the dissipation of smoke and fumes over larger areas, reducing the concentration from any particular source at ground level. This strategy for dealing with air pollutants in general prevailed into the middle of the twentieth century.

Today, electric utility plants account for about 70% of annual SO_2 emissions and 30% of NO_x emissions in the United States. Mobile sources (transportation) also contribute significantly to NO_x emissions. More than 22 Tg (terragrams = 1 million metric tons) of SO_2 are emitted into the atmosphere each year in the United States, and 180 Tg are emitted globally.

A. SO_x

SO_2 is the principal form of anthropogenic sulfur emission and it is released primarily by combustion of fossil fuels. SO_2 dissolves in water droplets where it can be oxidized to H_2SO_4. This has a low vapor pressure and tends to form aerosol particles. These aerosols can form salts with Ca^{2+}, Mg^{2+}, or NH_4^+ and can become nuclei for the condensation of water and formation of clouds.

The residence time of sulfur in the atmosphere is controlled by the processes that deposit it to the ground. About half of the sulfur burden of the atmosphere is removed by dry deposition, although the ratio of dry to wet deposition varies widely.

The total amount of sulfur emitted into Earth's atmo-

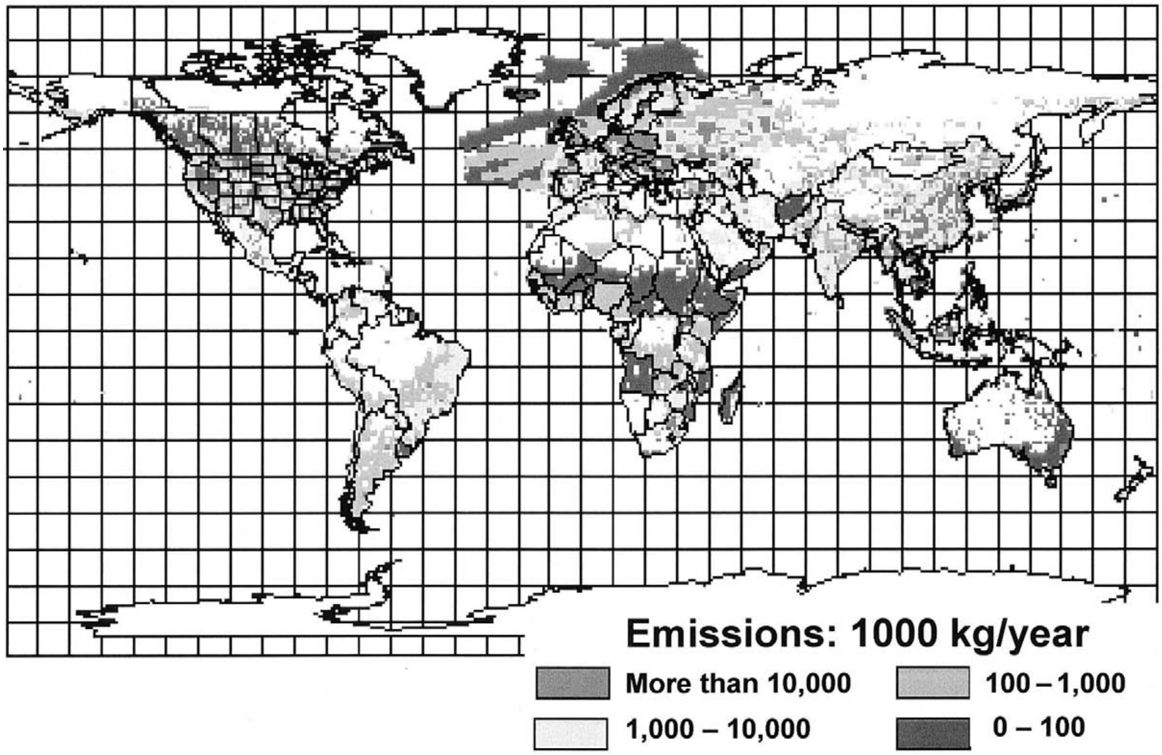


FIGURE 1 Annual sulfur oxide emissions as sulfur on a 1° × 1° latitude/longitude grid (1000 kg/year) (Canadian Global Emissions Interpretation Centre, a joint initiative of Canadian ORTECH Environmental, Inc., and Environment Canada). See also color insert, this volume.

sphere in 1985 (the reference year) was 90 Tg (calculated as elemental sulfur, equivalent to 180 Tg of SO_2) from all sources (Fig. 1). By 1990, global anthropogenic emission of sulfur was 85 Tg (170 Tg as SO_2). Emissions of SO_2 in the United States peaked in 1977 at 32 Tg. By 1985, U.S. emissions of SO_2 had declined to 25 Tg (Table I). The largest source of SO_2 is electric power plants, accounting for 69% of U.S. SO_2 emissions. More than 90% of these power plant emissions are from combustion of coal.

TABLE I

U.S. Sources of SO_2 and NO_x Emissions to the Atmosphere in 1985 in Tg per Year[a]

Source	SO_2	NO[a]
Electric utilities	14.6	6.15
Nonutility combustion	2.4	2.98
Nonferrous smelters	0.6	
Residential/commercial	0.6	0.64
Other industrial processes	2.1	0.63
Transportation	0.8	7.61
Miscellaneous		0.20
Total	21.1	18.21

[a] From NAPAP (1990).

Natural sources of sulfur emissions globally contribute as much as 7% of total sulfur emissions. Dimethyl sulfide released from the oceans is oxidized in the atmosphere to sulfate and may account for 60% of these natural emissions. Volcanism (20%), decomposition processes in soils and plants (15%), and coastal wetlands (3%) are other sources. In eastern North America and northern Europe and Britain, natural sources of sulfur emissions are of little importance as sources of SO_x and NO_x, accounting for less than 1% of regional sulfur emissions according to Environmental Protection Agency (EPA) studies.

B. NO_x

Human activities have more than doubled the emissions of fixed nitrogen to the atmosphere, surpassing the total of all natural sources. The primary form emitted by fuel combustion is NO_2. The largest single anthropogenic source is the transportation sector (40%), with fossilfired utilities (30%) a close second. Anthropogenic emissions of NO_x have risen more or less steadily from about 3 Tg released in the Year 1900. By 1985, 36–41 Tg of NO_x was emitted globally (Fig. 2), with more than 22 Tg emitted in the United States alone. About 30% was from electric utilities and 43% from the transportation sector in that year. Natural sources of NO_x to the atmosphere, which may contribute to the formation of NO_3^-, are less well-known. There are natural sources of NO_x emissions in soil, lightning, and stratospheric injections that account for 6, 5, and 0.1%, respectively, of the total of anthropogenic plus natural sources of NO_x emissions in the United States according to NAPAP reports (1990).

NO_x chemistry is complex and leads to the formation of nitric acid (HNO_3). Nitric acid gas can react with aerosols such as sea salt, adsorb onto particles of soil, or react with ammonia to form NH_4NO_3. Ammonia is emitted to the atmosphere in urban and agricultural areas largely due to human activities. The rate at which nitric acid is deposited from the atmosphere as a dry gas is much faster than the deposition velocity of NH_4NO_3; thus, the presence of ammonia facilitates the long-range transport of NO_x.

III. PRECIPITATION CHEMISTRY

Wet deposition is relatively easy to collect and to evaluate. Most of the wet-deposited pollutants arrive in just a few major precipitation events. Dry deposition is a slower and more continuous process, but it is quite difficult to measure and local factors that alter wind turbulence and seasonal factors are important to the accuracy of measurements. On a regional basis, wet and dry deposition are approximately equal, but in urban areas or near to major emission zones dry deposition may be considerably greater than wet deposition.

Precipitation samples are collected in areas of the world that are remote form sources of SO_x, such as Point Barrow in Alaska, Mauna Loa in Hawaii, and at the South Pole, by the Global Trends Network (GTN). In such remote areas, the average pH of precipitation is closer to 5.0 than to 5.6, which is the pH value that might be expected from an equilibrium of atmospheric CO_2 in pure water. Apparently, natural sources of acidity (e.g., oceanic or wetland emissions of sulfur) reduce pH below this expected value. It is also clear, however, that anthropogenic pollutants, SO_4^{2-} and NO_3^-, contrib-

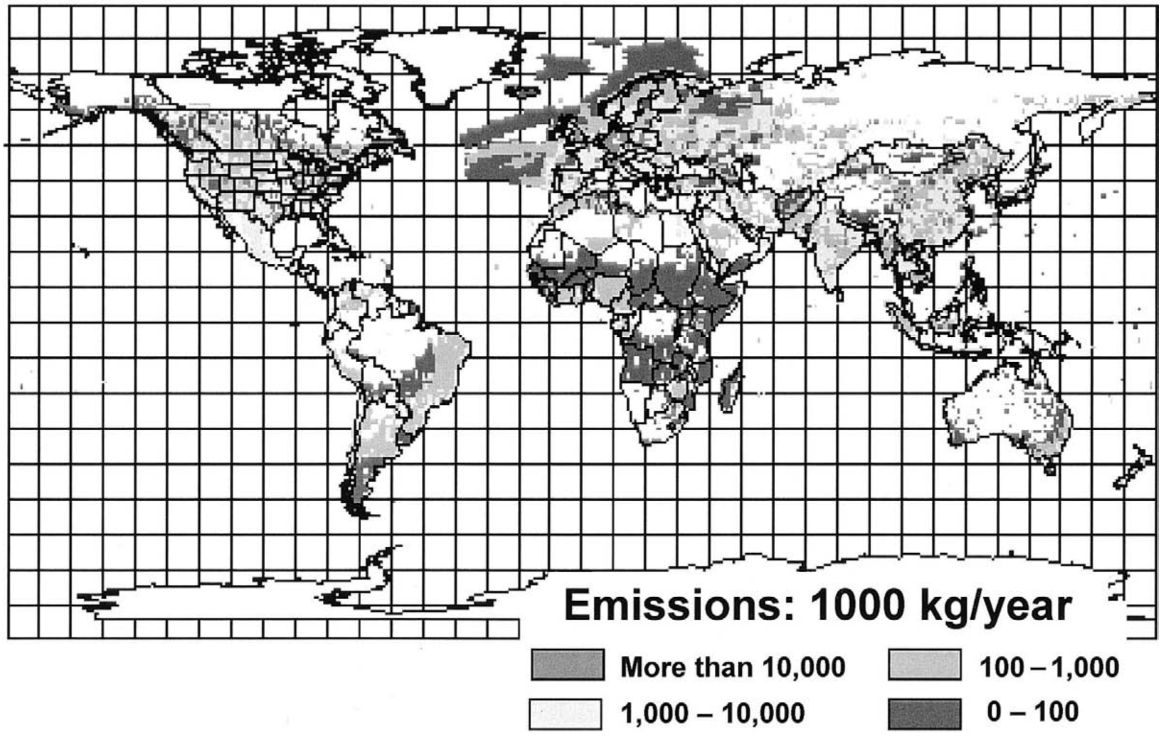


FIGURE 2 Annual nitrogen oxide emissions as nitrogen on a 1° × 1° latitude/longitude grid (1000 kg/year) (Canadian Global Emissions Interpretation Centre, a joint initiative of Canadian ORTECH Environmental, Inc., and Environment Canada). See also color insert, this volume.

ute to this acidification and no area of the world is free of anthropogenic pollutants.

Eastern North America and northern Europe are receptor regions downwind from large area sources. Compared to the remote regions, these receptor regions receive nine times more SO_4^{2-}, 14 times more NO_3^-, seven times as much NH_4^+, and six times as much H^+. The sources of these contaminants are the upwind emissions from industrial and population centers. NAPAP (1990) reports that all forms of precipitation over much of eastern North America, on average, are quite acidic (Fig. 3). Mean annual "wet" precipitation (weighted by the volume of each precipitation event) was in the range pH 5.0–4. Individual rain episodes with pH near 3.0 are observed in the northeastern United States. There is great variability in the amount deposited across continental areas. For example, the average annual deposition of sulfur species (Fig. 4) at Argonne, Illinois (in 1985–1987), was 23.6 kg/ha, whereas at Pawnee, Colorado, it was 1.7 kg/ha. In northern Europe, including Britain, all of Scandinavia in the north and down to mid-France and northern Italy, and east to the border of Russia, the annual average pH of precipitation was below 4.9 in 1985.

HNO_3 is deposited as a dry gas from the atmosphere onto vegetation or other moist surfaces. The rate of HNO_3 deposition is much faster than the deposition velocity of the nitrate aerosol, ammonium nitrate (NH_4NO_3). Thus, the presence of ammonia facilitates the long-range transport of NO_x (Fig. 5).

IV. EFFECTS

A. Forests

There are numerous examples of forest dieback related to local sources of pollution. For example, SO_2 emissions at near-ground level from a copper smelter in Sudbury, Ontario, killed forests, grasses, and soil organisms and created a local landscape that some called a moonscape. Similar situations exist around point emission sources elsewhere. Acid deposition, however, is a problem associated with the long-range transport of

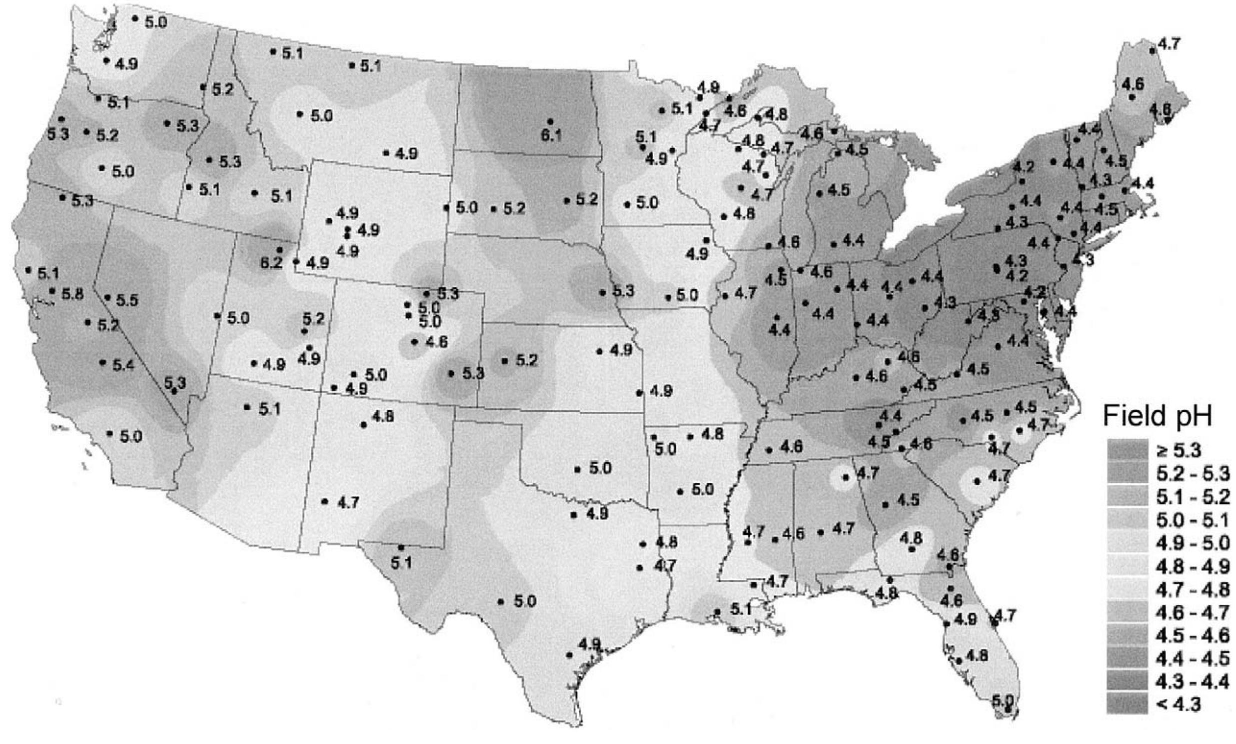


FIGURE 3 Annual average pH of precipitation in the United States for 1998. Samples were analyzed at the Central Analytical Laboratory, National Atmospheric Deposition Program/National Trends Network (reproduced with permission from the National Atmospheric Deposition Program/National Trends Network, 1998).

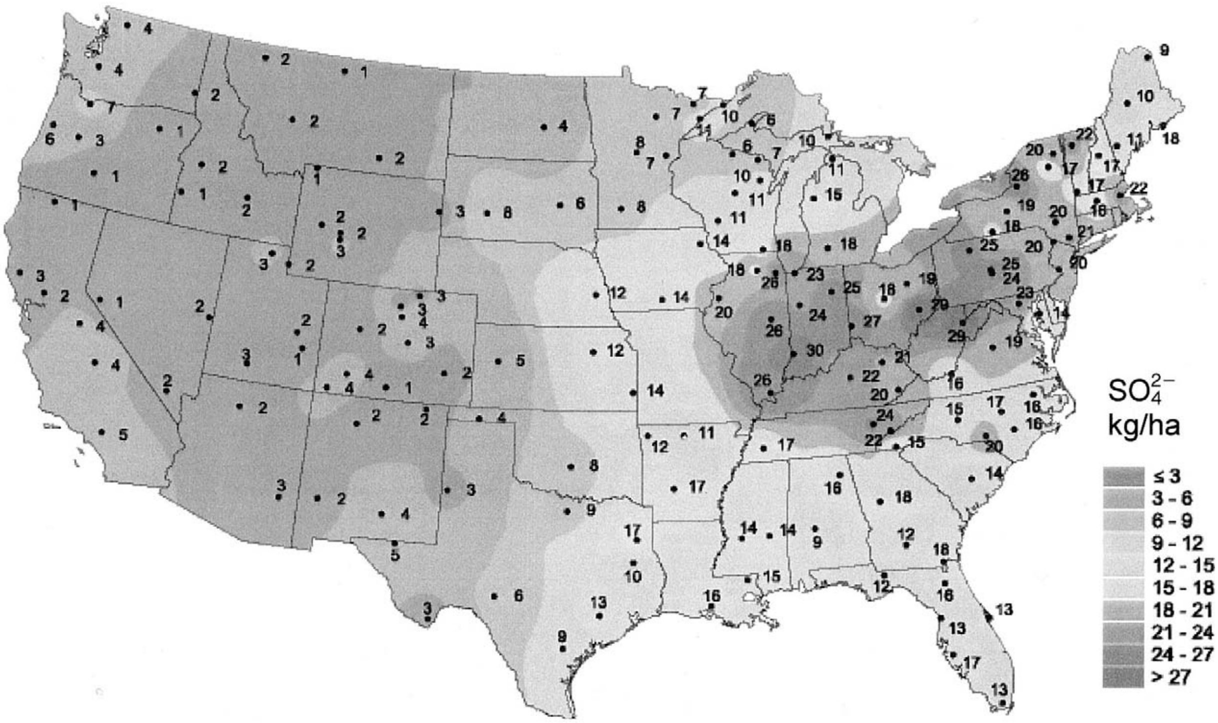


FIGURE 4 Estimated sulfate deposition in the United States in 1998 (reproduced with permission from the National Atmospheric Deposition Program/National Trends Network, 1998).

pollutants, with receptor areas hundreds or even thousands of kilometers from the emission sources.

Throughout Europe, forests are plagued by thinning of the topmost branches, called "crown-thinning." In Eastern Europe, where high-sulfur coal has been consumed in prodigious quantities, areas in which high concentrations of atmospheric pollutants (especially SO_2) occur have undergone significant forest injury and dieback, although the extent of damage is not well quantified. In the western part of Germany, many forest declines appeared to be occurring in the 1970s that were attributed in the popular press to acid rain. The term "Waldsterben" (forest death) was used to describe the situation. This led to increased public pressure for environmental protection in general and for research on topics relating to acid deposition in particular. Large-scale surveys of forest condition were carried out in Germany where there were many regional declines. The overall conclusion of the surveys, however, was that less than 20% of western German forest area was classified as damaged and no large-scale deforestation was occurring. In fact, it was found that the rate of forest stress seemed to be decreasing and surveys showed that Norway spruce (*Picea abies*) injury was reported to be 9% less in 1988 than in 1985.

Partly due to forest problems in Germany, it was suspected that forest productivity and the health of forest in North American ecosystems might be compromised by acidification, acting either directly on vegetation or through changes in forest soils. In Canada, 45% of the land is covered with forests, as are 33% of the United States and 21% of Mexico. NAPAP organized a Forest Response Program in 1985 to address issues of forest damage in general and the role that acid deposition might play in such damage. Similar research activities were carried out in other countries, including Canada and Norway, in which forests cover extensive areas in regions most heavily impacted by acid deposition.

It is known that forests are impacted by a variety of stresses and it is often difficult to isolate specific causes of local forest decline. Fire, insect pests, microbial infestations, poor management practices, and even natural aging can act alone or together (Table II) in causing forest decline. Severe forest declines have occurred in the past. For example, during the period 1871–1885 an estimated 50% of mature spruce trees in the Adirondack Mountains died from unknown causes. Another example is that of "fir waves," in which patches of balsam fir (*Abies balsamea*) in the Appalachian region die out

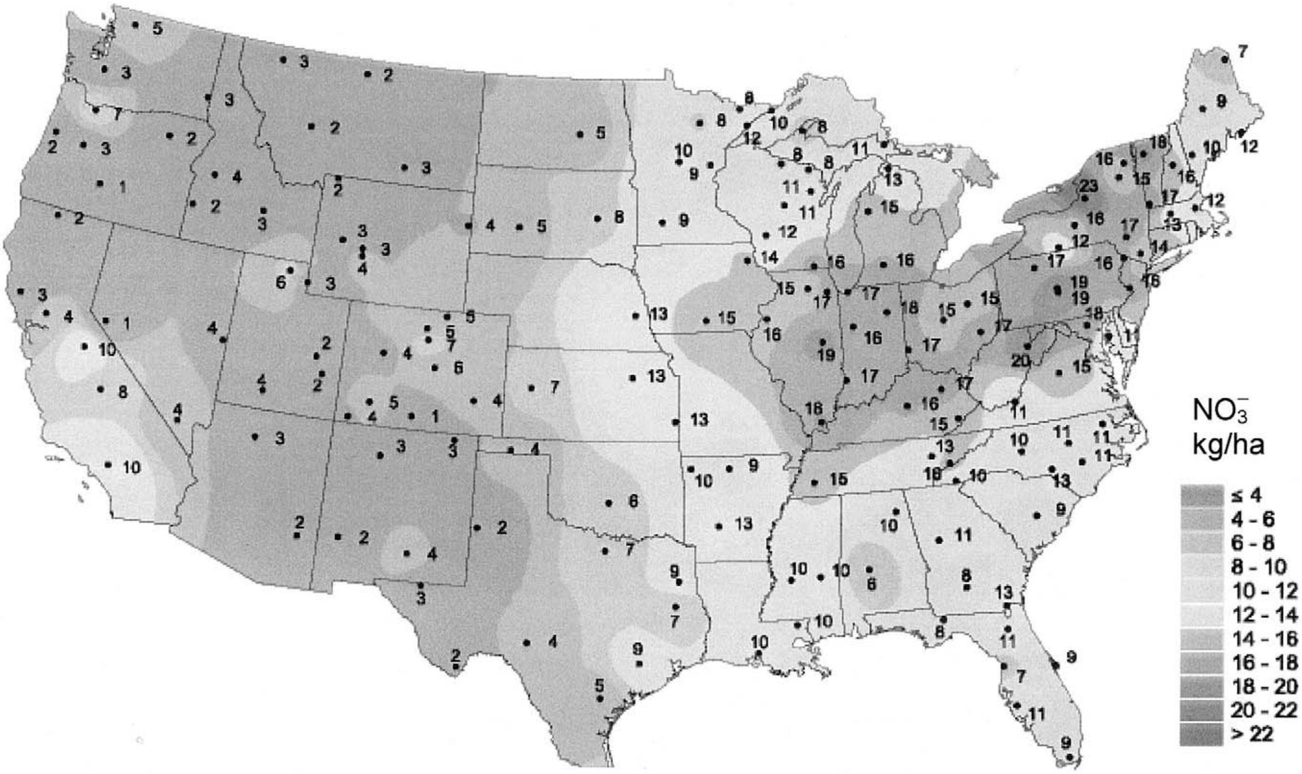


FIGURE 5 Estimated nitrate ion deposition in the United States in 1998 (reproduced with permission from the National Atmospheric Deposition Program/National Trends Network, 1998).

in a wave-like pattern across the landscape. It was thought that a potentially stressing condition, such as heavy loading of a region by acid deposition, might cause a general weakened condition that makes forests more susceptible to other problems (Barnard *et al.*, 1990).

In the United States red spruce (*Picea rubins Sarg.*), high-elevation spruce trees that populate the ridges of

TABLE II

Major Natural Stresses in U.S. Forests and Their Effects in 1985[a]

Stress	Affected area (10^6 ha)	Nature of impact
Wildfire	1.7	$750 million damage to timber resource in 1985
Gypsy moth	0.7	Tree defoliation concentrated in southern New England and mid-Atlantic states
Mountain pine beetle	>1.6	Mortality often severe in infested stands
Southern pine beetle		
Spruce bud worm	2.1	Tree defoliation and mortality concentrated in Maine
Western spruce bud worm	5.1	Tree defoliation and mortality, primarily in the Rockies, intermountain region, and the Pacific Northwest
Fusiform rust	6.1	$35 million in lost timber value through volume and quality reduction in southern pines.
Dwarf mistletoe	9.0	Annual timber volume loss of 10.8 million m^3 in the West
Root diseases	>6.0	Annual timber value loss of 6.8 million m^3 in the West; substantial but unquantified impact in other regions

[a] From NAPAP (1990).

the Appalachian Mountains from Maine to Georgia, has undergone a period of dieback of 25–50% from the 1950s through 1989. The dieback is associated with severe winter injury that kills the terminal and lateral shoots, and the repetition of this injury can lead to overall stress, susceptibility to injury from fungi and insects, reduced growth, and tree death. NAPAP concluded that acid deposition has contributed to this dieback, but the mechanisms of injury are uncertain.

Many experimental studies of acid deposition effects on trees have been conducted, including exposure to SO_2, ammonium-sulfate aerosols, artificial acid rain, or acid rain plus ozone. In general, these experimental studies did not show that a significant negative effect that might stress forests or reduce forest growth was caused by pollutant levels similar to those associated with acid deposition. In reviewing all the evidence concerning acid deposition effects on forests, NAPAP concluded that (i) most forests in eastern North America are exposed to acidic deposition and to elevated concentrations of ozone but do not show signs of unusual growth loss or tree decline; (ii) spatial and temporal patterns of tree health and productivity are not consistently related to estimated levels of pollutant exposure; and (iii) except for red spruce at high elevations, there is no general deterioration in the health or productivity of eastern forests and no consistent relationship between forest health and atmospheric deposition. It was also concluded, however, that there were indications of stress to forests that should be monitored carefully, including the following: (i) Ambient ozone levels are affecting plant physiology in some species, (ii) acid deposition and ozone can interact in tree injury, and (iii) alteration of forest element cycles may affect species composition.

B. Crops

Scientists have been studying the effects of air pollutants on plants for many years. Research in this area accelerated after the discovery of ozone as a constituent in the ambient air that is toxic to plants. (Effects of exposure of plants to elevated ozone, although related to the issue of acid rain, are not included in the context of this article.) There are many examples of plant injury due to acute fumigation by air pollutants from local sources. Whereas there is convincing experimental evidence that acid deposition can damage crop plants, reports of crop loss due to acid deposition (excluding local point sources of fumigation) are as scarce as hens' teeth. Since the mid-1960s, more than 5000 research reports have been published dealing with the topic, including acute and chronic exposures of plants to SO_2, various forms of NO_x, and elevated H^+ as well as other contaminants. This extensive research effort is probably more a reflection of the importance of crops to both human nutrition and the agricultural economics than to any observation of crop loss due to long-range transport of pollutants. The conclusions reached by NAPAP (1990) are as follows: (i) Conditions capable of causing acute injury to vegetation as a result of exposure to present-day levels of gaseous air pollutants (including ozone) are rare, occur only during unusual conditions of atmospheric stability, and are confined to a limited number of areas, and (ii) acute injury to vegetation due to acidic deposition is virtually unknown.

C. Soils

The characteristics of natural terrestrial ecosystems are largely determined by the properties of their soils. In northern forest soils there is a layer of humus containing exchangeable bases overlying the mineral soil. When H^+ is deposited with acid deposition into the forest, or generated by growth of vegetation, the humus layer and soil minerals can retain this H^+ and release an equivalent amount of base-forming cations (Ca^{2+} and Mg^{2+}) that generate alkalinity. In this way, much of the acid input may be neutralized for as long as the exchange capacity lasts. In sufficiently moist areas precipitation can leach base-forming cations as fast as the rate of primary mineral weathering of these ions, causing soils to become acidic due to natural processes. As soil acidity increases, aluminum (Al) becomes increasingly soluble but also adsorbs onto clay minerals, and soil solution Al can hydrolyze to increase soil H^+ concentrations.

Acid deposition will have an acidifying effect or will leach base cations (Ca^{2+}, Mg^{2+}, K^+, and Na^+) or both in soils with low cation exchange capacity (CEC). In areas where sulfate adsorption in soils is low, such as in the Adirondack Mountains, sulfur deposited from the atmosphere behaves more or less conservatively and passes through soils into lakes and streams as SO_4^{2-}. However, the total amount of strong acid anion must be balanced by an equivalent amount of cations. If CEC is depleted, then Al^{3+} is mobilized by H^+ exchange. Al^{3+} and some of the H^+ associated with acid rain enter the runoff water, thus maintaining the charge balance, and the water is acidified.

Recent studies by the U.S. Geological Survey (1999b) and others have found that calcium is being depleted from forest soils in the eastern United States as a consequence of both acid deposition and uptake by roots.

In agricultural soils, agronomic practices of tilling, fertilization, and liming are far more important factors in altering soil chemistry than acid deposition. Additions of HNO_3 by acid deposition even may be beneficial for forest growth since nitrogen is frequently a growth-limiting nutrient though benefits from H_2SO_4 deposition are viewed as minimal.

D. Ground-water

The total atmospheric load of acids exceeds the ability of soils to provide bases in many areas of Europe. This is especially true in areas underlain by slow-weathering, base-deficient rocks, such as granite, gneiss, quartzite, and sandstone. In these areas with continued acid deposition loading, base saturation (the fraction of CEC occupied by exchangeable base cations) can be expected to decrease steadily and eventually approach zero. This will result in acidification of ground-water.

Norway is highly impacted by acid deposition, with average precipitation pH in the range 4.3–4.5. Ground-water in many areas is unusually acidic, with pH in the range 5.2–5.7. Such water is quite corrosive for copper pipes. In Denmark, the pH of deep well water decreased from 6.5 to 5.6 between the 1950s and 1980s. The Hartz Mountains of Germany also receive very acidic precipitation. Sulfate concentrations in ground-water there have risen from 5 mg /liter in the 1960s to a current value near 20 mg/liter and the water from 33 springs has high concentrations of metals (cadmium, 0.1–2.0 μg/liter; zinc, 50–150 μg/liter; and nickel, 5–20 μg/liter). In some cases, spring-water pH is less than 4.

E. Surface Waters

Acidification of surface waters is defined as a decrease in alkalinity, or acid neutralizing capacity (ANC). As acids are added to water, the H^+ increases and ANC is reduced. The most significant impact of acid deposition is that on surface waters, in which it causes acidification and ecological damage in many thousands of lakes and streams. In some sensitive waters, fish species such as brook trout have been completely eradicated. This is one of the few environmental impacts that have been clearly demonstrated for "acid rain", and it is politically important because it has resulted in the loss of fisheries and recreational value, which people can readily understand.

Unpolluted surface waters sensitive to acidification generally are found to be in the range pH 6–7 with low ANC. Watersheds with significant amounts of carbonate minerals can readily buffer inputs of acid by creating alkalinity. Watersheds with soils low in minerals with Ca^{2+} or Mg^{2+} have little ability to generate ANC. When unpolluted waters have ANC < 100 microequivalents per liter (μeq/liter) they are classified as sensitive to acidification.

Most surface water acidification is due to the deposition of sulfate that provides a long-term and rather steady base-load of strong acid anions. Nitrate is important in the episodic acidification associated with particular precipitation events or snowmelt, which can dramatically increase acidity of lakes and streams. In "brown" waters, the concentration of acidic, humic materials is high and also contributes to acidity. However, acidification of surface waters—that is, the change in chemistry over time in many "sensitive" areas of the world—is clearly the result of excessive SO_4^{2-} concentrations due to acid deposition and not a consequence of the presence of natural organic acids.

Several factors interact to make the waters or a region susceptible to acidification due to inputs of strong acids. The most important of these are (i) proximity to emissions sources, (ii) regional meteorological patterns, (ii) bedrock geology, and (iv) topography (NAPAP, 1990). The most heavily impacted regions are located downwind from large emissions sources in Great Britain and northern Europe and of the central industrial region in the United States and Canada. The importance of geographic location and wind direction is illustrated by the fact that the very sensitive waters of northern Minnesota and southwestern Ontario are not yet as severely impacted as are the waters of the Northeast and maritime provinces. There is evidence that acidification is occurring in this region. Acid deposition is emerging as a significant problem in Asia, but there is scant information available on actual or potential ecological consequences there.

1. Surface Water Surveys

Large-scale, statistically based surveys of lakes and streams have been conducted in several countries to evaluate actual and potential impacts of acid deposition on surface waters. In most cases, these surveys were designed to investigate waters in regions thought to be sensitive to acidification because of high rates of acid deposition or in regions having waters with low ANC. In these sensitive areas, concentration of SO_4^{2-} is strongly correlated to wet SO_4^{2-} deposition. The U.S. National Surface Water Survey found that all of the sampled lakes and streams with $\text{pH} < 5.5$ or $\text{ANC} < 0$ occur in areas receiving precipitation with $\text{pH} < 5.0$ and wet SO_4^{2-} deposition loading greater than 10 kg/ha/year. Furthermore, acidic lakes in which SO_4^{2-} is the dominant anion are not found in regions receiving wet SO_4^{2-} deposition less than 10 kg/ha/year. The Norwegian national acid deposition effects project found that in southern Norway, where acid deposition is great and soils are both thin and base deficient, 75% of the lakes are acidic and SO_4^{2-} from acid deposition is the dominant anion.

A survey of 8506 lakes was carried out in 10 regions of Canada and 56% of these lakes were found to be sensitive to acidification ($\text{ANC} < 100$ μeq/liter). In some areas, up to 84% of the lakes were found to be sensitive and as many as 60% of the lakes in some areas were very sensitive ($\text{ANC} < 50$ μeq/liter). Acidic lakes, those with $\text{ANC} < 0$ μeq/liter, comprised 5% of all lakes in the sample, and up to 24% in one region were acidic.

Nitrogen as well as sulfur deposition can contribute to chronic and episodic acidification of surface waters. Unlike SO_4^{2-}, however, NO_3^- is usually conserved within watersheds because of plant uptake of N. Exceptions to this rule, however, are seen in those areas of the world in which NO_3^- deposition is unusually great. In streams of southwestern Norway, NO_3^- concentrations exceed 10 μeq/liter and nitrate can make up over 10% of strong acid anions.

Trends in precipitation and stream-water chemistry were examined at eight precipitation monitoring stations during the period 1984–1996 by the U.S. Geological Survey (1999a). In the northeastern United States, results indicate that decreases in atmospheric deposition of SO_4^{2-} have resulted in decreased precipitation acidity.

2. Episodic Acidification

Episodic acidification of surface waters occurs as a consequence of acidic snowmelt and acidic rain events. Snow accumulation is one mechanism by which strong acid anions may be stored and concentrated within watersheds receiving acid deposition. Accumulated contaminants in the winter snowpack can be released at the onset of melting so that 50–80% of the SO_4^{2-} received over a period of months may be released from the snow-pack with the first 30% of the meltwater. Thus, early snowmelt runoff waters in areas such as southern Norway and the Adirondack Mountains carry pollutant loads that are greatly elevated. Regions with heavy snowfall can be especially susceptible if the rate of acid deposition is high.

F. Marine Waters

In some areas, the amount of nitrogen in soils, from agricultural fertilizers and acid deposition, exceeds the needs of vegetation and NO_3^- is discharged in surface waters. In Scandinavia, NO_3^- accumulated in the snowpack is discharged so quickly by melt-water that it is not taken up by vegetation. In such cases, rivers carry NO_3^- to estuaries and bays. Acid deposition also falls directly on marine waters, increasing the loading of nitrate. Chesapeake Bay, for example, receives 30–50% of its nitrogen from acid deposition and this contributes to eutrophication. In Scandinavia, acid deposition has contributed to excessive nitrogen in marine waters that appears to cause phytoplankton blooms.

G. Aquatic Biota

Acid deposition, by acidifying surface waters, causes widespread ecological damage (Table III). There is a widespread misconception, however, that acidified lakes and streams are "dead." The fact is that even the most acidified surface waters have many organisms. Species of protozoa and insects are found at pH 2.0, rotifers and Cladocera occur at pH 3.0, and even some fish are found at pH 3.5. Acidified waters are not "dead;" they can be full of life—but this is life run amok in ecosystems severely out of balance.

1. Microbial Communities

Abnormal accumulations of coarse organic matter are observed on the bottoms of some acidified lakes and dense felt-like mats of fungal hyphae can cover much of the bottom areas. The accumulation of debris and fungal mats both seal off the mineral sediments from interaction with the overlying water and hold organically bound nutrients that would have become mineralized and available if normal decomposition had occurred. Reductions in nutrient availability may have a negative feedback effect on microorganisms, further inhibiting their activities. Acidification can also inhibit microbial nitrogen cycle activities. Reduction of microdecomposer activities may also have a direct effect on invertebrates feeding on microbial biomass associated with decomposing litter, further inhibiting litter removal and nutrient regeneration. Bacteria respond to acidification gradually, with no clearly delineated thresholds above pH 5.5. Treatment of lakes with lime raises pH and causes rapid decomposition of the organic debris and fungal mat and increases in bacteria in the water, indicating that microbial communities were inhibited at low pH.

TABLE III

Summary of Biological Changes Due to Surface Water Acidification[a]

pH range	General biological effect
6.5–6.0	Small decrease in species richness of phytoplankton, zooplankton, and benthic invertebrate communities resulting from the loss of a few highly acid-sensitive species, but no measurable change in total community abundance or production
6.0–5.5	Loss of sensitive species of minnows and dace, such as blacknose dace and fathead minnow; in some waters decreased reproductive success of lake trout and walleye, which are important sport fish species in some areas.
	Visual accumulation of filamentous green algae in the littoral zone of many lakes and in some streams
	Distinct decreases in the species richness and change in species composition of the phytoplankton, zooplankton, and benthic invertebrate communities, although little if any change in total community biomass or production
	Loss of many common invertebrate species from the zooplankton and benthic communities, including zooplankton species such as *Diaptomus silicis, Mysis relicta,* and *Epsichyura lacustris*; many species of snails, clams, mayflies, and amphipods and some crayfish
5.5–5.0	Loss of several important species of fish, including lake trout, walleye, rainbow trout, and smallmouth bass, as well as additional nongame species such as creek chub
	Further increase in the extent and abundance of filamentous green algae in lake littoral areas and in streams
	Continued shifts in species composition and decline in species richness of the phytoplankton, periphyton, zooplankton, and benthic invertebrate communities; decreases in the total abundance and biomass of benthic invertebrates and zooplankton may occur in some waters
	Loss of several additional invertebrate species common in oligotrophic waters, including *Daphnia galeata mendotae, Diaphanosoma leuchtenbergianum,* and *Asplanchna priodonta;* all snails, most species of clams, and many species of mayflies; stoneflies, and other benthic invertebrates
	Inhibition of nitrification
5.0–4.5	Loss of most species of fish, including most important sport fish species such as brook trout and Altantic salmon; few fish species able to survive and reproduce below pH 4.5
	Measurable decline in the whole-system rates of decomposition of some forms of organic mattter, potentially resulting in decreased rates of nutrient cycling
	Substantial decrease in the number of species of zooplankton and periphyton communities; measurable decreases in the total community biomass of zooplankton and benthic invertebrates in most waters
	Loss of zooplankton species such as *Tropocyclops prasinus mexicanus, Leptodora kindtii*, and *Conochilis unicornis*; and benthic invertebrate species including all clams and many insects and crustaceans
	Reproductive failure of some acid-sensitive species of amphibians such as spotted salamanders, Jefferson salamanders, and the leopard frog

[a] From NAPAP (1990).

2. Aquatic Plants

Freshwater ecosystems are supported by photosynthesis within the water body and by inputs of organic debris from the surrounding land. Primary production, the synthesis of living material from inorganic elements by photosynthesis, is carried out in freshwaters by a wide variety of plants, including leafy macrophytes, mosses, and algae.

i. Phytoplankton

Phytoplankton are microscopic plants that live suspended in the water column. Phytoplankton communities are usually quite diverse, with typically several dozens of species. Evidence concerning the impact of acid deposition on phytoplankton comes from the large synoptic lake surveys in North America and Europe (particularly Scandinavia), from experiments in which the chemistry of lakes was changed intentionally to evaluate acidification impacts, and from studies of artificially enclosed “mesocosms” in which variables such as pH and nutrient concentration can be manipulated and the species composition controlled. These studies demonstrate that decreasing pH lowers species richness and diversity. Simplification of phytoplankton communities is especially acute over the range of pH 6–5.

ii. Periphyton

Periphyton is the material growing on submerged surfaces in freshwaters. It is dominated by microalgae that often form long filaments or sheets that can cover the sediments, plants, or other objects in water. The Periphyton can become a complex community of algae, bacteria, fungi, and a variety of invertebrates.Periphyton species richness decreases with increasing acidity.

A striking phenomenon is the proliferation of attached algae in both streams and lakes, with increasing acidity. Common water macrophytes, such as *Lobilia dortmana* and *Isoetis lacustris,* are festooned with filamentous algae and the bottoms of acidic streams may be covered with attached algae. Such increases in algal mass occur despite reduced specific rates of photosynthesis, indicating factors other than a preference for low pH are allowing algae to accumulate. Several ecological factors appear to contribute to algal proliferation at sub-optimal pH: decreased microbial activity, reduced competition among algal species allowing only the most acid tolerant to proliferate, and reduced grazing by invertebrates.

iii. Macrophytes

Aquatic macrophytes, including the vascular plants, mosses, liverworts, and macro-algae, are important elements of aquatic ecosystems. Macrophytes help to stabilize sediments and shorelines, form breeding grounds for some fish and many invertebrate species, and are a food source for waterfowl and mammals such as beavers and moose. Swedish limnologist Ole Grahn and colleagues (1974, 1977) studied acidification in Swedish lakes. Acid deposition decreased pH from 5.6 to 4.8 and brought about a regression of communities including *Lobilia,* whereas communities dominated by the aquatic moss *Sphagnum* expanded from average coverage of about 8% to cover half of the littoral zones in a period of just 6 years. *Sphagnum* has a significant ion exchange capacity that results in the sequestration of Ca^{2+} and Mg^{2+}, thus withdrawing cations from the water. The extensive moss mats covered much of the lake bottom and reduced both mineralization and exchange between the sediments and the overlying water. Large mats of *Sphagnum* are infrequently observed in North American lakes. Changes in macrophyte communities in acidified lakes may also be associated with other chemical changes, such as the availability of Ca^{2+}. Raising the pH of lakes and increasing the Ca^{2+} supply by liming dramatically reduced *Sphagnum* communities.

3. Invertebrates

Lakes and streams that are not impacted by acidification have a diverse set of invertebrates with many species of insects, worms, crustaceans, and mollusks. In clear, unpolluted streams with moderate alkalinity in the pH range 6–8, there may be 70–90 species, of which a few are plentiful. As pH decreases below 5.7–5.4, so do the numbers of species. Mayflies, caddis-flies, crustaceans, and mollusks become rare or even disappear from the community. Changes in other elements of the ecosystem can alter its food supply, and changes in the faunal community may increase or decrease predation on a particular invertebrate species. There are critical pH thresholds below which survival of a particular species is greatly reduced. Not only the acidity of the water but also the concentrations of beneficial elements such as calcium and potassium and the concentration of toxic metals, particularly dissolved aluminum, are critical features in the responses of invertebrates to acidification and can greatly influence the rate of mortality at low pH.

i. Zooplankton

Zooplankton are small (normally less than 2 mm long) aquatic invertebrates, including copepods, cladocerans (water fleas), and rotifers, living in the water columns of lakes or slow-moving streams. Some are herbivores grazing on phytoplankton and some are predatory carnivores, and they are an important food source to fish and waterfowl. Synoptic surveys of hundreds of surface waters in Scandinavia and North America found that the number of zooplankton species in a water sample is highly correlated to pH. Several species of Cladocera and Rotifera are seen to increase in abundance with decreasing pH. Thus, zooplankton density (animals per liter of water) is not as sensitive to pH as is species richness since the more tolerant species can increase in number to replace missing species. In some acidified lakes there is a shift toward large-bodied zooplankton predators that may be due to decreased predation by fish, with the fish having been excluded due to acidification. This increases predation on smaller zooplankton. There is ample evidence that population-level changes are linked to increasing concentrations of Al^{3+} and reproductive failure.

ii. Macroinvertebrates

The aquatic macroinvertebrates are normally highly diverse assemblages of organisms. They are ecologically important to healthy ecosystems, assisting in the breakdown of litter and detritus, as grazers of algae, as predators of other invertebrates and juvenile stages of fish, and as a food source to fish and water-fowl. Surveys of macroinvertebrates in hundreds of lakes and streams in areas receiving large inputs of acid deposition clearly show that species richness declines sharply with increasing acidity. Several species of mayflies, amphipods, crayfish, and virtually all snails and clams are quite sensitive to low pH and are lost from the fauna of acidified waters. Species richness, diversity, and biomass decrease with decreasing pH. This is evident even in the pH range 7.0–6.0. A few species are very tolerant of both low pH and elevated aluminum concentration.

In acidified lakes, fish predation is reduced or eliminated altogether by the disappearance of the fish and the acid-tolerant and predatory water boatmen and backswimmers (Hemiptera) may become important predators of other invertebrates. The amphipod *Gammarus lacustris* is absent from waters with pH lower than 6.0. Acidification experiments show that the progression through larval stages of *Lepidurus* arcticus is retarded with increasing pH and toxicity is complete at pH 5.5

Impacts of acid deposition on lake ecosystem have been studied experimentally by David Schindler and colleagues at Canada's Experimental Lakes Area. They intentionally acidified whole lakes over a period of several years from near neutral to pH near 5.0. Changes in macroinvertebrate communities became apparent even as pH changed from 6.8 to 5.9. Species numbers were reduced and others became more abundant as pH continued to decrease. At pH 6.0–5.8, the freshwater shrimp *Mysis relicta* became extinct. At pH 5.1, the crayfish *Orconectis virilis* became extinct, apparently due to a combination of factors including the inability to calcify their shells, reproductive failure, and direct toxicity to juveniles.

Stoney streams normally contain a rich assemblage of macroinvertebrates. When ANC is moderate and pH is approximately 6 or higher, there may be 70–90 taxa present. When stream acidity is lower than 5.5, many of these taxa are scarce or absent. Mayflies, some caddis flies, mollusks, and crustaceans are the most sensitive. The fauna is impoverished by acidification and may contain only half the numbers of taxa found in unacidified soft-water streams.

Experimental acidification of streams has demonstrated detrimental impacts on macroinvertebrates including reduced numbers of species. Some intolerant species drift downstream to avoid the acidified waters and in this way can be eliminated from the acidified stream reach. In headwater streams, in which acidification is most severe, re-colonization would be unlikely.

4. Fish

In 1926, fisheries biologists noted that there was a widespread reduction in the catch of salmon in the major rivers of southern Norway, and in 1959 acid deposition was identified to be the cause. In seven rivers (mean pH 5.1) of this impacted region, 150 metric tons of Atlantic salmon were taken in 1900. Atlantic salmon were virtually eliminated from these rivers due to acidification, despite efforts to improve the fishery through hatcheries and stocking. Meanwhile, the catch from 68 other Norwegian rivers (mean pH 6.6) in areas not subjected to such intense acid deposition increased from 155 metric tons in 1900 to nearly 300 metric tons.

In 10 rivers of Nova Scotia, Canada (mean pH $<$ 5.5 in 1980), where angling catch of Atlantic salmon was good in the mid-1930s, the catch went to zero by the 1980s. In rivers in the area that are less impacted by acidification (1980 mean pH $>$ 5.0) the catch in 1980 was about the same as in the mid-1930s.

Lake fisheries are also severely impacted by acidification. In southern Norway, by the mid-1970s brown trout disappeared from half of the lakes in which they formerly occurred. By 1980, 30% of the remaining brown trout populations and 12% of the perch populations disappeared from the region. Lakes from which fish populations were lost had lower pH, higher concentrations of aluminum, and lower concentrations of calcium. Many lakes in the region have been studied intensively and fish kills associated with episodic acidification during acidic rain events and snowmelt are observed in some of the lakes in which fish stocks are declining.

Surveys in many areas show a strong relationship between species richness and lake pH, including lakes in Norway, Sweden, The Netherlands, Scotland, the LaCloche Mountains of Ontario, the Adirondack Mountains of New York, northern Wisconsin, and the Upper Peninsula of Michigan.

Acidification problems in the United States and Canada may be greater than is indicated by large-scale surveys because they tend to miss episodic acidification events. Lakes and streams throughout North America, including high-elevation lakes in the West, experience such events. Many have low ANC and are therefore sensitive to acidification. Episodic acidification causes fish kills and can severely damage entire year-classes of fish. In the Adirondack Mountains 70% of all sensitive lakes are at risk of episodic acidification. In the mid-Appalachian region, 30% of sensitive streams, or seven times the number of chronically acidic streams, can become acidified by such episodes.

Both low pH and elevated Al^{3+} concentrations are known to cause these impacts through the loss of the ability to regulate body salts and leakage of salts through the gills. Recruitment failure, due to effects on all stages from egg to adult, is an important mechanism for the loss of populations of fish.

5. Other Animals

Much less is known about the impact of acid deposition on other animals, such as amphibians (frogs and newts), birds, and mammals. Many species of amphibians are declining throughout the world, but the causes are not obvious and large-scale species declines have not been clearly linked to acid deposition. Fish-eating birds are impacted by losses of fish populations. Elevated aluminum concentrations are associated with decreased reproduction in passerine birds. The concentration of cadmium is elevated in the internal organs of large herbivores in areas of North America and Scandinavia where surface water acidification is a problem. Elevated concentrations of mercury in fish in these areas may lead to contamination of otters and mink.

H. Materials

The problem of corrosion due to air pollutants has been known for 150 years or more. Angus Smith (1852) noted,

> "The presence of free sulfuric acid in the air sufficiently explains the fading of colours in prints and dried goods, the rusting of metals, and the rotting of blinds. It has been observed that the lower portions of projecting stones in buildings were more apt to crumble away than the upper; as the rain falls down and lodges there and by degrees evaporates, the acid will be left and the action on the stone be much increased."

Acid deposition contributes to corrosion of many types of materials, including painted surfaces, metals and carbonate stone (limestone and marble), masonry, carbon steel, zinc, nickel, and some paints and plastics. Both wet and dry deposition participate in the corrosion process. This is particularly a problem for limestone and marble buildings and monuments throughout the world. Monuments and buildings, such as the Taj Mahal, have suffered extensive damage. The great Gothic churches, such as the Cologne Cathedral and Notre Dame in Paris, as well as more ancient structures such as the Coliseum in Rome are melting away. Many structures that have withstood normal weathering processes for 1000 years or more are, in recent times, suffering extensive damage, as are newer buildings such as the U.S. Capital Building in Washington, DC.

I. Health

Sulfur dioxide can have serious health impacts on people. Persons with asthma can experience difficulty in breathing when exposed to SO_2 while exercising for as little as 5 minutes. Studies of air pollution episodes in London and New York in the mid-twentieth century found that among the elderly, the very young, and those with pre-existing respiratory disease, increased mortality followed exposure to average ambient SO_2 concentrations of >0.19 ppm for 24 hours. In other epidemiological studies, the U.S. EPA found that persons living within 20 km of large point sources of SO_2 emissions were at risk from such episodes. Lowering sulfate aerosol levels will reduce the incidence and the severity of asthma and bronchitis. Reductions in NO_x and O_3 emissions are also expected to have a beneficial impact on health effects. The Clean Air Act and subsequent amendments resulted in reductions of SO_2 emissions. Consequently, air quality has improved. Nevertheless, approximately 46 million people in the northeaster United States continue to be exposed to air quality that does not meet EPA's health-based air standards for one or more of the six criteria pollutants.

J. Visibility

Emissions that cause acid rain also reduce transparency of the atmosphere and decrease atmospheric visibility. The aesthetic properties of outdoor scenery in parklands such as the Shenandoah and the Great Smoky Mountains are noticeably reduced by hazy air. Particles with diameters less than 2.5 μm, dominated by sulfate and ammonium in eastern North America, account for 75–95% of visibility reduction. In the western United States, the sulfate contribution is less: 20–50% in rural areas and 10–20% in urban areas. A measure of atmospheric visibility is the visual range, which is the distance over which one can see. In the U.S. Southwest, the median value is about 150 km. On the U.S. Pacific and Atlantic coasts the median visual range is 20–50 km. Summertime haziness has generally increased in the eastern United States since the late 1940s, and this is largely due to increased sulfate aerosols. The trend is not uniform, however; haziness increased most in the Southeast.

This has been a gradual process so that most people think that a slightly whitish haze on a clear, sunny day is normal. This haziness is what Stephen Schwartz has called the "white house effect" and it is a consequence of sulfate aerosols in the size range of 0.1–1 μm diameter. The sulfate is nearly all from oxidation of SO_2 emitted by fuel combustion. These aerosols act as condensation nuclei for water and the formation of clouds. Under some conditions aerosols may be reduced to 1% of their usual concentration by convective upward movement of air, cloud formation, particle scavenging, and precipitation. Such conditions make the air unusually dry. On these rare days the sky seems unusually blue, and this is a hint of what our ancestors could see on most sunny days.

V. REGULATION

An early attempt to limit precursors of acid deposition globally was the 1979 Geneva Convention on Long-Range Transboundary Air Pollution. This established emissions limits for sulfur and nitrogen that have in general been met. In the late 1990s emissions of SO_x in Europe were approaching half the amount emitted in the 1970s. In the United States, the Clean Air Act and subsequent amendments have brought about large reductions in SO_2 emissions. By 1996 the annual wet SO_4^{2-} deposition over much of the eastern United States declined by 10–25%. Despite progress in reducing emission of SO_2 in North America and Europe, the global problem of acid deposition is not likely to disappear. In Asia, emissions of SO_2 are expected to triple in the period 1990–2010.

See Also the Following Articles

AIR POLLUTION • ATMOSPHERIC GASES • FOREST ECOLOGY

Bibliography

Binkley, D., Driscoll, C. T., Allen, H. L., Schoeneberber, P., and McAvoy, D. (1989). Acid Deposition and Forest Soils, *Ecological Studies* Vol. **72**. Springer-Verlag, New York.

Downing, R., Ramankutty, R., and Shah, J. (1997). *RAINS-ASIA: An Assessment Model for Acid Deposition in Asia,* pp. 11, 48, 54, and Table 3 (p. 27). World Bank, Washington, DC.

Grahn, O., Hultberg, H., and Lander, L. (1974). Oligotrophication—A self-accelerating process in lakes subjected to excessive supply of acid substances. *Ambio* **3**, 93–94.

Henriksen, A., and Brakke, D. F. (1988). Increasing contributions of nitrogen to the acidity of surface waters in Norway. *Water Air Soil Pollut.* **42**, 183–201.

National Acid Precipitation Assessment Program (NAPAP) (1990). *Acidic Deposition State of Science and Technology Vol. II. Aquatic Process and Effects* (P. M. Irving, E. T. Smith, and N. B. Clancy, Eds.). NAPAP, Washington, DC.

National Atmospheric Deposition Program (NRSP-3)/National Trends Network (1998). NADP Program Office, Illinois State Water Survey, 2204 Griffith Dr., Champaign,IL 61820.

Schindler, D. W., Mills, K. H., Mally, D. F., Findlay, D. L., Shearer, J. A., Davies, I. J., Truner, M. A., Linsey, G. A., and Cruikshank, D. R. (1985). Long-term ecosystem stress: The effects of years of experimental acidification on a small lake. *Science* **228**, 1395–1401

Schwartz, S. E. (1996). The white house effect—Shortwave radiative forcing of climate by anthropogenic aerosols: An overview. *J. Aerosol. Sci.* **27**, 359–382.

Sutcliffe, D. W., and Hildrew, A. G. (1989). Invertebrate communities in acid streams. In *Acid Toxicity and Aquatic Animals* (R. Morris, E. W. Taylor, D. J. A. Brown, and J. A. Brown, Eds.), pp. 13–29. Cambridge Univ. Press, New York.

U.S. Geological Survey (USGS) (1996). *Trends in Precipitation Chemistry in the United States, 1983–94: An Analysis of the Effects in 1995 of Phase I of the CAAA of 1990, Title IV*, USGS 96–0346. USGS, Washington, DC.

U.S. Geological Survey (USGS) (1999a). *Trends in Precipitation and Stream-Water Chemistry in the Northeastern United States* (D. W. Clow and M. A. Mast, Eds.), USGS monograph *(ftp://bqsnt.cr.usgs.gov/manilles/Clowfact3.pdf)*. USGS, Washington, DC.

U.S. Geological Survey (USGS) (1999b). *Soil and Calcium Depletion Linked to Acid Rain and Forest Growth in the Eastern United States* (G. B. Lawrence and T. G. Huntington, Eds.), USGS monograph *(ftp://bqsnt.cr.usgs.gov/manilles/WRIR984267.pdf)*. USGS, Washington, DC.

ADAPTATION

Michael R. Rose
University of California, Irvine

GLOSSARY

adaptation One or two of the following: a beneficial construct produced by an omnipotent being, the process of change established by natural selection, and a biological character that gives increased Darwinian fitness.

adaptationism The doctrine that all important evolutionary processes are dominated by natural selection, and that all significant biological characters increase an organism's fitness.

biological altruism Behavior of an organism such that the fitness of another organism is increased while its own fitness is decreased.

clutch size The number of eggs a bird lays in its nest at one time.

epistasis Interactions between genes at different chromosomal locations in the determination of phenotypic character values.

fitness Net reproductive output, discounted for any lack of viability.

genetic drift Accidents of segregation and recombination causing evolutionary genetic change.

group selection Selection between different populations or sub-populations based on attributes of the entire group, where these attributes usually are either selected against or not favored at the level of individual selection.

heterozygote An individual having two different alleles at a genetic locus.

hominid A great ape from the lineages most closely related to humans, where this may be a lineage ancestral to humans.

inbreeding The mating of close biological relatives.

individual selection Selection driven by differences in the net reproduction of individual organisms.

industrial melanism Selection for darker pigmentation as a result of industrial pollution, particularly in moths and butterflies.

linkage disequilibrium Nonrandom association of alleles on chromosomes.

meiotic drive Preferential segregation of a parasitic gene during gamete production.

phenotype The manifest biological character(s) of a particular organism.

phylogenetic Pertaining to evolutionary ancestry.

recombination The shuffling of gene combinations in the production of gametes, possibly by the physical breaking and rejoining of pieces of chromosomes.

Encyclopedia of Biodiversity, Volume 1

segregation Allocation of genetic variants ("alleles") to different gametes during sexual reproduction.

teleology The imputation of goal-directed behavior or structures.

ADAPTATION consists of one or two of the following: a beneficial construct produced by an omnipotent being, the process of change established by natural selection, and a biological character that gives increased Darwinian fitness.

I. HISTORICAL INTRODUCTION

A. Classical Times

The concept of adaptation is older than any scientific concept of evolution, and certainly older than Darwin's theory of natural selection. The founder of academic biology, Aristotle, gave adaptive explanations for many of the features of the living and nonliving world. Thus, the webbed feet of a frog can be said to be "for" efficient swimming, and thus they can be explained as an illustration of the universe being well made. This type of reasoning was commonplace in classical culture, which often assumed some type of benign natural order.

It is significant, however, that adaptive reasoning had its critics even in classical times. Lucretius, one of the most important of classical proto-scientists, was scathing about the wholesale imputation of function to body parts, even when such inferences were regarded as "common sense." Such criticisms of the concept of adaptation have waxed and waned ever since.

B. Pre-Darwinian Christendom

Biblical theology gave arguments about adaptation a new cast. The assumption that there was a single, benign, omnipotent Creator made the existence of well-constructed organisms a natural assumption. From the beneficence of the Creator, each organism must have been given the specific characteristics best suited to its role in the Creation as a whole. Indeed, this concept of benign, and efficient, creation was extended to physics, especially by Isaac Newton, an avid believer. The orbits of the planets were thereby interpreted as evidence of some type of adaptation, or suitedness, to a divine plan.

This universal adaptation gave rise to some interesting paradoxes for pre-Darwinian scientists. Did the Creator adapt organisms to the physical universe or was the physical universe created to fit the organisms? Was life on Earth based primarily on the chemistry of water and organic molecules because that was the biochemistry that could work on this particular planet? Or was the planet constructed by the Creator to fit the biochemistry that He already had in mind? How could such questions ever be resolved?

C. The Darwinian Theory of Adaptation

Darwin's theory of evolution by natural selection provided a natural solution to the two problems of what adaptations were and how they occurred. In Darwin's theory, selection operating on heritable variaon increased the frequency of individuals bearing attributes, "adaptations," which gave them increased fitness. Fitness in turn was to be defined as the net reproductive rate of individual organisms in the original version of Darwinism. Thus, Darwin's theory proposed that the environment drove the evolution of adaptations by determining the pattern of selection imposed on organisms.

Darwin's theory had many novel features for the biology of his time. First, it involved no omnipotent Creator beneficently organizing the arrangements of life. Second, there were no inner drives or teleologies shaping the process of organic change to an adaptive end, unlike the scheme of Lamarck and others, who were more influenced by Aristotle than was Darwin. Third, there was no overarching pattern to Darwin's process of evolution, and therefore adaptations might occur higgeldy-piggeldy, whenever selection made mortality or reproduction hinge on a particular attribute. All these features made Darwin's new biology of adaptation distasteful to many of the older generation of Victorian biologists, who were highly teleological in their thinking when not avidly creationist.

II. TWO COMMON DEFINITIONS

The nature of Darwinian theory instilled a large degree of ambiguity in the term adaptation. For a creationist, there is no process of adaptation, only what the Creator made and its beneficent nature. For the Darwinian, like the Lamarckian, there is necessarily a process of adaptation—a process by which adaptation is brought about. Then there is the product of adaptation as a process, which is called an adaptation as well.

There has been controversy regarding which of these basic meanings of the term is the true one or the correct one. However, we can follow Ernst Mayr, or Karl Pop-per, and reject the need to find any essentially true definition of adaptation. Instead, the two alternative definitions can both be used as appropriate.

A. Adaptation as a Process

The concept of adaptation as a process derives from the theory of natural selection. Therefore, a deeper consideration of this incarnation of adaptation requires study of natural selection. One of the basic intuitive expectations that most evolutionary biologists have is that natural selection should lead to the evolution of increased Darwinian fitness. However, this is not universally true. Mutation, segregation, recombination, meiotic drive, and frequency-dependent selection can force natural selection to produce a decrease in fitness.

Usually, this decrease in fitness is temporary. For example, if the heterozygote, at a locus with two alleles, is the most fit genotype, then if the population is initially composed entirely of heterozygotes segregation will cause an immediate decrease in fitness. However, this effect is confined to the first generation. Subsequent generations will have increasing or stable mean fitness as natural selection brings the population to the stable gene frequency equilibrium, at which mean fitness will also be at a local maximum. Analogous processes can occur with other genetic processes, such as recombination. Again, in some cases, mean fitness does not continue to decline.

However, there are also cases in which fitness may continue to decrease, and natural selection never produces a recovery in mean fitness. Meiotic drive is a well-known example. Meiotic drive occurs when some genes pervert segregation rates in their favor so that, despite being deleterious, they spread through populations. Another situation in which mean fitness can decrease is brought about by natural selection. The fitness of a mating with males and females can be a nonlinear function of the genotypes of the two organisms mating. If such nonlinearities are sufficiently severe, natural selection on fertility can actually drive fitness to increasingly lower levels, at least in theory. No prominent empirical examples of this process are currently known, however.

The point that these examples serve to make is that although it is conventional in evolutionary biology to expect improved adaptation from the action of natural selection, there is no absolute warrant, either in theory or in fact, for this assumption. Theory and experiment both indicate that a process of adaptation is usually brought about by natural selection. However, it is not always brought about by natural selection.

When natural selection does act, however, to establish a process of adaptation, what can we say about that process? This is one of the three major research projects of evolutionary biology (the other two being the inference of phylogeny and the study of the genetic material used by evolution). Thus, our understanding of adaptation as a process is undergoing continual upgrading as our understanding of natural selection improves.

At the most basic level, however, there are some essential features of adaptation by natural selection that can be considered as well established. There is no general or consistent pattern to natural selection. Specific populations may undergo very intense selection for a short period of time. One of the best studied examples is the recent work on the evolution of Darwin's finches on the Galapagos, particularly the effects of drought on bill size (Grant, 1986). On the other hand, for most populations, it is usually very difficult to detect the action of natural selection. It is either too weak or too variable in direction (Abrahamson and Weis, 1997). Some of the cases in which natural selection can be readily detected as working in each generation to produce adaptation involve human disturbances—ecological events that are unlikely to reflect the evolutionary situation of populations that have been spared artificial disruption. The classic example of this scenario is the evolution of wing camouflage in the moths of industrial Europe in which natural selection was generated because soot blackened the tree trunks on which these moths rested, making the light-colored moths stand out against a black background. The very artificiality of this case, however, underscores the point that we do not normally find such cases of unequivocal selection when we study natural populations.

This leads to the next major point about adaptation as a process: It is difficult to detect. Therefore, adaptation as a process tends to be assumed by evolutionary biologists more than it is actually demonstrated. Also, the teasing out of the mechanistic particulars of adaptation as a process is almost never accomplished. This central problem has led to a pervasive weakness in the scientific analysis of adaptation as a process, with unfortunate consequences.

B. Adaptation as a Product of Evolution

The view of adaptation as a product of evolution does not logically require that it be a product of natural selection. An adaptation can arise evolutionarily from selection on some other character(s), or it might occur from some nonselective process, such as inbreeding or genetic drift. Thus, for example, a spider's web might have been evolved because of selection for prey capture, but it may also constitute an adaptation that enables spiders to obtain water from dew condensing on the web. This raises the following question: If adaptation is divorced from the process by which it arose, then how is it to be distinguished from the other characteristics of an organism?

The conventional solution to this problem is to define adaptations as those products of evolution, however generated, that enhance the fitness of the organism. Nominally, this requires that fitness be measured with and without the character(s) that is presumed an adaptation. This is a difficult enterprise for two reasons.

First, it is often difficult to perform the surgery, or other manipulation, required to make organisms without the adaptation in question. Recently, however, it is in precisely this area in which significant progress has been made in studies of adaptation (e.g., Sinervo and Basolo as cited in Rose and Lauder, 1996). Evolutionary biologists are now successfully ablating tissues and grafting on additional body parts in order to test the fitness consequences of the possession or loss of particular structures that are being evaluated for their status as adaptations. Manipulation of clutch size by removal or addition of eggs has been a traditional method in studies of vertebrate life history adaptations. Research in this area now manipulates fertility and egg size using a variety of techniques, including microsurgery. The resources of modern molecular biology are likely to give evolutionary research even more power to manipulate phenotypes.

Second, the measurement of fitness is difficult in most organisms. In organisms that reproduce strictly by dividing in two, without sex, fitness can be measured fairly easily from estimates of viability between bouts of fission. In every other kind of organism, sex and variable numbers of offspring make the estimation of fitness extremely difficult. Perhaps the worst character of all in the estimation of fitness is male mating success. This difficulty arises because the attribution of maternity is usually fairly secure, whereas the attribution of paternity is often pure speculation. This is an area in which the recent findings of behavioral ecology suggest considerable grounds for pessimism. Pairs of birds, for example, may indeed remain together for life, sharing the tasks of caring for young, foraging for food, and nest construction. However, molecular genetic analysis of pedigrees frequently reveals that the "monogamous" female has had sex with another male of the species, while the male has himself dallied. Similar patterns are well-known from human paternity cases. There are also species that are either highly promiscuous, such as chimpanzees, or ejaculate gametes externally, such as most fish. In these species, there are no mated pairs to keep track of over the long term. For these reasons, estimating the Darwinian fitness of an individual with a particular phenotype is often extremely difficult, if not practically impossible.

The fallback position of many biologists, especially functional morphologists, comparative physiologists, and behavioral ecologists, has been to use a surrogate for fitness. Such surrogates include mechanical efficiency, conservation of metabolic energy, and the number of copulations. The assumption is usually made that such surrogate measures will always be positively correlated with fitness. When they improve, fitness should increase. Unfortunately, it is precisely these characters that will show diminishing returns rather than a stable, positive correlation with fitness. Mechanical efficiency is patently not the only impact of structure on fitness. Structures may be costly to develop, or they may impede movement. Evolution is unlikely to maximize each and every "design feature" of an organism, even if there were no genetic constraints preventing the realization of any particular phenotype. Therefore, the expedient of using surrogates for fitness is not likely to be reliable in many cases.

If fitness cannot be accurately measured, and surrogates for fitness cannot be relied on, it is difficult to see how the concept of adaptation as a product of evolution can be used in most cases. There are pleas to the effect that some characters are so intuitively beneficial that they cannot reasonably be denied the status of adaptations. Legs must be adaptations for terrestrial locomotion, large brains must be adaptations for life as a tool user, and so on. However, limbs may be used for many functions, not just locomotion. The hominid brain has also been explained as an adaptation for social behavior, not the use of tools. Supposedly obvious cases become far from obvious once all possible scientific interpretations are taken into account.

III. EVIDENCE FOR ADAPTATION

If both basic definitions of adaptation are allowed, then there are two different lines of evidence for the existence of adaptation. The first is simply the action of natural selection. If adaptation is the process of natural selection, then any evidence for such selection is in turn evidence for adaptation. The second line of evidence is supplied whenever there are data showing an increase in fitness when a particular character is acquired. Together, the accumulated evidence bearing on both of these points helps establish the importance of adaptation as a feature and as an outcome of the evolutionary process.

A. Evidence for Natural Selection

If Darwin generally lacked evidence for natural selection in nature, modern evolutionary biology has supplied an abundance of such evidence (Endler, 1986), including classic studies of industrial melanism and recent studies of drought selection in Darwin's finches (Grant, 1986). However, there are many other examples of natural selection in the wild, dating back to W. R. F. Weldon's study of carapace width in estuarine crabs in the 1890s. Indeed, natural selection is such an obvious feature of the living world that it is now considered in discussions of such practical medical problems as the prescription of antibiotics and the treatment of the human immunodeficiency virus (Freeman and Herron, 1998). Thus, the general principle that there is a process of adaptation involving natural selection is not in any reasonable doubt.

The evidential problems instead concern the importance of the process in any particular instance. The idea of an adaptive process shaping the course of evolution is very attractive because it can be used to support the interpretation of evolutionary change in terms of natural selection. However, as discussed previously, the demonstration that such a process is occurring is usually very difficult. Also, the possibility that other evolutionary processes are involved—processes that do not involve adaptation by natural selection for the character of interest—cannot be dismissed out of hand. This renders most casual post hoc invocations of natural selection essentially dubious. Whatever the specific features of natural selection, casually invoking it as an explanation for all features of life is no longer reputable behavior in evolutionary biology.

This means that, although there are some specific studies that provide excellent evidence for adaptation by natural selection, in most cases scientists are not in a position to interpret an evolutionary process as being driven by natural selection. It may be allowed as a possibility, but further study is usually required before a particular evolutionary change can be considered as being brought about by natural selection, even when such an interpretation seems intuitively natural.

B. Evidence for Increased Fitness

Even if it is difficult to establish the nature of the evolutionary process, surely the products of evolution are easier to categorize as adaptive? For the reasons discussed previously, however, it is often difficult to make an accurate determination concerning whether or not the possession of a particular character increases fitness. In particular, it is not enough to show that a particular function (e.g., locomotion) has been improved, perhaps by a longer hind-limb, because such demonstrations do not define the effect on fitness as a whole. A particular function could be improved while fitness is reduced.

Currently, some of the best demonstrations of adaptation come from the field of behavioral ecology. Of particular value have been manipulative experiments which change the behavior or morphology of study animals and plants. These studies have supplied many instances in which artificially created deviants have demonstrably reduced fitness (Sinervo and Basolo as cited in Rose and Lauder, 1996).

There is much potential for the study of molecular biology to extend the power of manipulation in the study of adaptation, particularly with genetic transformation and the insertion of genes with artificially inducible expression. Some of these studies have measured the effects on adult survival of gene insertions (Fleming and Rose, 1996). Fitness could also be measured in such experiments.

An alternative approach is to measure the relationship between the variation of a character and fitness in polymorphic populations. Much of modern evolutionary quantitative genetics collects data of this kind. One of the central concerns in these studies is the delimitation of optima for fitness as a function of quantitative characters.

Finally, artificial selection can be used to generate perturbed values for selectable characters. If artificial selection is then relaxed, and the original character state was adaptive, natural selection should drive the character to its original state. This has been observed in only a few cases (Service *et al.*, 1988). Such patterns of reversion are expected to occur especially when there are trade-offs between functional character, such that high values of one character are associated with low values of other characters. This situation is particularly important for the use of surrogate measures for fitness.

IV. CRITIQUE OF ADAPTATIONISM

A. Adaptationism

The record of classical thinking, such as that of Aristotle, illustrates the extent to which the human mind is attracted to the idea of beneficent organization in the natural world. Many classical scholars believed this even when they had no particular scientific theories to buttress the concept.

Darwinian evolution is thus an almost irresistible temptation for those who wish to infer function in the living world. Darwinism guarantees a role for natural selection in evolution, and it guarantees the existence of adaptation among the characters of organisms. However, it does not guarantee that selection and adaptation must be everywhere prepotent, at all times, and in all respects.

Nonetheless, there is a variant form of Darwinism that flourished particularly in the 1950s and 1960s—a variant that assumed that all the attributes of an organism are shaped by natural selection to the end of increased fitness. In this version of Darwinism, now called "adaptationism," all characters are adaptations and all nontrivial evolutionary processes are driven by natural selection. In effect, this school of thought made the study of evolution tantamount to the study of adaptation.

Among the effects of adaptationism on scientific practice was the notion that there must always be an adaptive explanation for every organ, structure, or behavior. Therefore, if an adaptive explanation for a particular structure has not been found, greater efforts must be made to discover its adaptive value. Alternative evolutionary processes (genetic drift, inbreeding, meiotic drive, etc.) must not be considered until all possible adaptive explanations have been tried and found wanting.

During its heyday, adaptationism put adaptation at the center of evolutionary biology, and to some extent at the center of all biology. The many theoretical and experimental problems facing the study of adaptation were minimized or dismissed altogether.

B. The Rejection of Adaptationism

From the late 1960s until the early 1980s, adaptationism suffered a series of blows from which it has yet to recover. The first of these was the detection of a vast amount of molecular genetic variation, first by protein electrophoresis and later by DNA sequencing. The significance of this finding for adaptationism is that most species appear to have far more segregating genetic variation than is likely to be explicable in terms of natural selection. Therefore, natural selection probably is not prepotent at the molecular level. The current scientific consensus is that many of the alleles that arise and eventually become fixed during evolution are merely neutral variants of already extant alleles. A great deal of genetic evolution has occurred, but much of it has not been driven by natural selection.

A second event was the publication of *Adaptation and Natural Selection* by George C. Williams in 1966. One of the common evasions of the adaptationists was to invoke group selection when they could not explain a particular character in terms of individual selection. Thus, many of the social behaviors of colonially nesting birds were explained in terms of adaptations for group selection. Williams pointed out that, usually, these explanations were highly dubious. He argued that the inference or explanation of adaptations required greater restraint, particularly regarding social behavior. This undercut group selection, one of the ways in which adaptationists had been able to discover adaptations underlying seemingly maladaptive behavior, such as biological altruism. In so doing, Williams also helped expose the extent to which adaptationism was based more on dogma than on well-founded science.

The third, and culminating, event in the decline of adaptationism was the publication of the paper, "Spandrels of San Marco," by Stephen Gould and Richard Lewontin (1979). In this paper, Gould and Lewontin hold up for ridicule the adaptationist assumption that there is a history of selection for every significant attribute of an organism. They follow Voltaire in his satirizing of such intellectual figures as Spinoza, particularly their boundless belief that "this is the best of all possible worlds," except that Gould and Lewontin satirize the adaptationist assumption of an all-powerful beneficent natural selection.

These events essentially undermined adaptationism as a dominant movement within evolutionary biology. Adaptationists remain scattered throughout biology, including such fields as molecular biology, comparative physiology, and systematics. However, the powerful hold that they had on evolutionary biology in the 1950s was broken.

V. ADAPTATION AFTER ADAPTATIONISM

Although adaptationism was clearly in error with regard to the universality of adaptation in the living world, its deposition brought with it an overreaction. Many evolutionary biologists effectively rejected the concept of adaptation as a whole. They refused to work on the problem and they criticized those who did. Since 1980, the study of phylogeny has become the central concern of evolutionary biology. For some, the study of adaptation is now a marginal, somewhat disgraceful, practice within biology as a whole.

Replacing the study of adaptation was the study of "constraints." Constraints in evolutionary biology are factors that prevent the achievement of an optimal adaptive outcome. Constraints have been discovered promiscuously in the evolutionary machinery: lack of genetic variation, linkage disequilibrium, too much environmental variation, too little environmental variation, epistasis, temporally variable selection, spatially variable selection, and so on. Evolutionary biology went from a doctrine in which adaptation was everywhere to a doctrine in which adaptation had disappeared to be replaced by paralyzing constraints.

In the 1990s, there was some stabilization of views on the topic of adaptation. Less of a pariah among evolutionary topics, an edited volume titled *Adaptation* was published by Rose and Lauder in 1996. Evolutionary biologists were spending more time using experimental and other techniques that could test for adaptation rather than simply assuming its presence or absence.

The comparative study of adaptation was greatly improved by an infusion of phylogenetic techniques. For example, if it is hypothesized that the gill structure of a fish species is an adaptation to a new way of life in salt water, but two species that had evolved in fresh water exclusively also have this gill structure, then the phylogenetic information indicates that the basic adaptive hypothesis is not correct.

Laboratory selection is currently used more often to study adaptation, with greater replication and greater attention to designs that can be used to make inferences about selection. The great advantage of performing selection in laboratories is that selective processes can be studied with greater statistical power and control, compared to the "experiments" of nature, all of which are unique and uncontrolled. For example, instead of studying the water physiology of two desert insect species compared to that of two forest insect species, evolutionary biologists select insects under conditions of desiccation using replicated selection lines and controls maintained free of desiccation (Bradley *et al.*, 1999). Instead of dealing with possible historical accidents that might have differentiated species in the wild, selected laboratory populations provide good material for critically testing theories of adaptation to particular environmental conditions.

Another major development in the study of adaptation has been the use of natural populations, especially manipulated natural populations, in studies that approximate laboratory experiments. Reznick and Travis (cited in Rose and Lauder, 1996) have studied guppy evolution in the streams of Trinidad. Multiple streams pass in parallel through highly uniform drainage systems, giving the streams isolated guppy populations—populations that evolve under effectively identical conditions. This experimental system has provided tremendous opportunities for the study of adaptation in the wild with both replication and controls.

The study of adaptation now proceeds with much more skepticism than in the past. Simultaneously, empirical methods have been greatly improved. The prospects have never been brighter for a genuine scientific analysis of adaptation, as opposed to the blithe speculations of the past.

See Also the Following Articles

DARWIN, CHARLES • EVOLUTION, THEORY OF • INBREEDING AND OUTBREEDING • PHENOTYPE, A HISTORICAL PERSPECTIVE • PHYLOGENY • RECOMBINATION

Bibliography

Abrahamson, W. G., and Weis, A. E. (1997). *Evolutionary Ecology across Three Trophic Levels: Goldenrods, Gallmakers and Natural Enemies*. Princeton Univ. Press, Princeton, NJ.

Bradley, T. J., Williams, A. E., and Rose, M. R. (1999). Phyiological responses to selection for desiccation resistance in *Drosophila melanogaster*. *Am. Zool.* **39**, 337–345.

Endler, J. A. (1986). *Natural Selection in the Wild*. Princeton Univ. Press, Princeton, NJ.

Fleming, J. G., and Rose, M. R. (1996). Genetics of aging in *Drosophila*. In *Handbook of the Biology of Aging* (E. L. Schneider and J. W. Rowe, Eds.). Academic Press, New York.

Freeman, S., and Herron, J. C. (1998). *Evolutionary Analysis*. Prentice Hall, Upper Saddle River, NJ.

Gould, S. J., and Lewontin, R. C. (1979). The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc. R. Soc. B* **205**, 581–598.

Grant, P. R. (1986). *The Ecology and Evolution of Darwin's Finches*. Princeton Univ. Press, Princeton, NJ.

Rose, M. R., and Lauder, G. V. (Eds.) (1996). *Adaptation*. Academic Press, New York.

Service, P. M., Hutchinson, E. W., and Rose, M. R. (1988). Multiple genetic mechanisms for the evolution of senescence. *Evolution* **42**, 708–716.

Williams, G. C. (1966). *Adaptation and Natural Selection*. Princeton Univ. Press, Princeton, NJ.

ADAPTIVE RADIATION

Rosemary G. Gillespie,[*] Francis G. Howarth,[†] and George K. Roderick[*]
**University of California, Berkeley and †Bishop Museum*

GLOSSARY

adaptive shift A change in the nature of a trait (morphology, ecology, or behavior) that enhances survival and/or reproduction in an ecological environment different from that originally occupied.

allopatric speciation The process of genetic divergence between geographically separated populations leading to distinct species.

character displacement Divergence in a morphological character between two species when their distributions coincide in the same ecological environment compared to overlap of the character in question in the two species when they are geographically separated.

convergence The evolution of similar characters in genetically unrelated or distantly related species, often as the result of selection in response to similar environmental pressures.

ecological release Expansion of habitat, or ecological environment, often resulting from release of species from competition.

founder effect Random genetic sampling in which only a few "founders" derived from a large population initiate a new population. Since these founders carry only a small fraction of the parental population's genetic variability, radically different gene frequencies can become established in the new colony.

key innovation A trait that increases the efficiency with which a resource is used and can thus allow entry into a new ecological zone.

natural selection The differential survival and/or reproduction of classes of entities that differ in one or more hereditary characteristics.

sexual selection Selection that acts directly on mating success through direct competition between members of one sex for mates or through choices made between the two sexes or through a combination of both modes.

sympatric speciation The process of genetic divergence between populations occupying the same geographic range leading to distinct species.

taxon cycle The repetitive pattern by which widespread dispersive stage I populations or species give rise to more restricted and specialized stage II populations or species; subsequent divergence leads to stage III local endemics.

Numerous definitions of adaptive radiation have been proposed. Almost all incorporate the idea of diversification in ecological roles, although they differ in their emphasis on relative rates of proliferation. Here, we propose a definition that seeks to be general but at the same time removes any implication of process: Adaptive radiation is a pattern of species diversification in which different species within a lineage occupy a diversity of ecological roles, with associated adaptations.

I. HISTORY OF THE CONCEPT

Beginning with the work of Darwin (1859) on the Galapagos fauna, the concept of adaptive radiation, in terms of diversification of ecological roles by means of natural selection, has been recognized. The term was first used by Osborn (1902) in describing parallel adaptations and convergence of species groups on different landmasses. Subsequently, it was developed as a major tenet for arguments presented in the modern synthesis by Huxley (1942). Simpson (1953), working on paleontological data, discussed the importance of key innovations in triggering adaptive radiation. For a detailed history of the concept of adaptive radiation, see Givnish (1997). Much recent information has been added, particularly during the past decade with the rise of molecular methods (Givnish and Sytsma, 1997).

II. NONADAPTIVE RADIATIONS

The term "nonadaptive radiation" has been used to describe situations in which species proliferation has not been attended by diversification of ecological roles (Gittenberger, 1991). When proliferation is simply a consequence of isolation, with isolated sibling species maintaining similar ecological affinities, then the radiation cannot be considered "adaptive." As will be described later, isolation has been invoked to explain the initial divergence of taxa in some radiations (e.g., Galapagos finches and cichlid fish), with the adaptive phase not occurring until recently diverged sibling species become sympatric. However, there are some cases of nonadaptive radiation, with many allopatric and ecologically similar species. Most of these radiations are caused by changes in topography that, instead of opening up new habitats, have served simply to isolate a previously more widespread species. For example, isolated mountaintops and other continental refugia have allowed species long periods of evolution in isolation, without any ecological change. This may lead to patterns of considerable genetic distance between morphologically similar species from different isolates (Schneider and Moritz, 1999). Similarly, diversification of snails on islands has frequently been attributed to topographical isolation [e.g., Crete (Gittenberger 1991) and Medeira (Cameron *et al.*, 1996)]. In general, it appears that (i) nonadaptive radiation occurs if there is isolation without any novel ecological opportunity and (ii) coexistence of species within a lineage will not occur in nonadaptive radiations but is a primary characteristic of adaptive radiations.

III. FACTORS UNDERLYING ADAPTIVE RADIATION

The common requirement for triggering adaptive radiation is the opening up of ecological space. This may be allowed by intrinsic factors, i.e., something that changes in the organism to allow radiation to occur; for example, evolution of tolerance toward noxious plant chemicals (Farrell and Mitter, 1994; Mitter *et al.*, 1988). Alternatively, it may occur as a result of extrinsic factors; for example, it has been reported to occur in geological history after an influx of nutrients into the system (Vermeij, 1995), in recent evolutionary time when new islands are colonized (Wagner and Funk, 1995; Liebherr and Polhemus, 1997), and in ecological time when a new habitat opens (Rainey and Travisano, 1998).

For ancient radiations, it is often difficult to determine the relative importance of intrinsic and extrinsic factors in allowing adaptive radiation. Factors associated with such radiations include (i) coincidence (after a slight delay) with major extinction episodes (Sloan *et al.*, 1986) and (ii) radiation of a group frequently starting from a small, unimpressive set of species from an earlier period. For example, fossil ammonites (shelled cephalopod mollusks) reveal episodes of tremendous proliferation and extinction through the Devonian, Triassic, Jurassic, and Cretaceous (Fig. 1; Lehmann, 1981). Echinoderms show a similar pattern, originating in the Ordovician and undergoing small radiations until all but one lineage went extinct by the end of the Permian (Smith, 1984). These then radiated extensively in the Triassic–early Jurassic, and the current diversity of forms remains similar to what arose at that time.

The great placental radiation (>4300 species) has

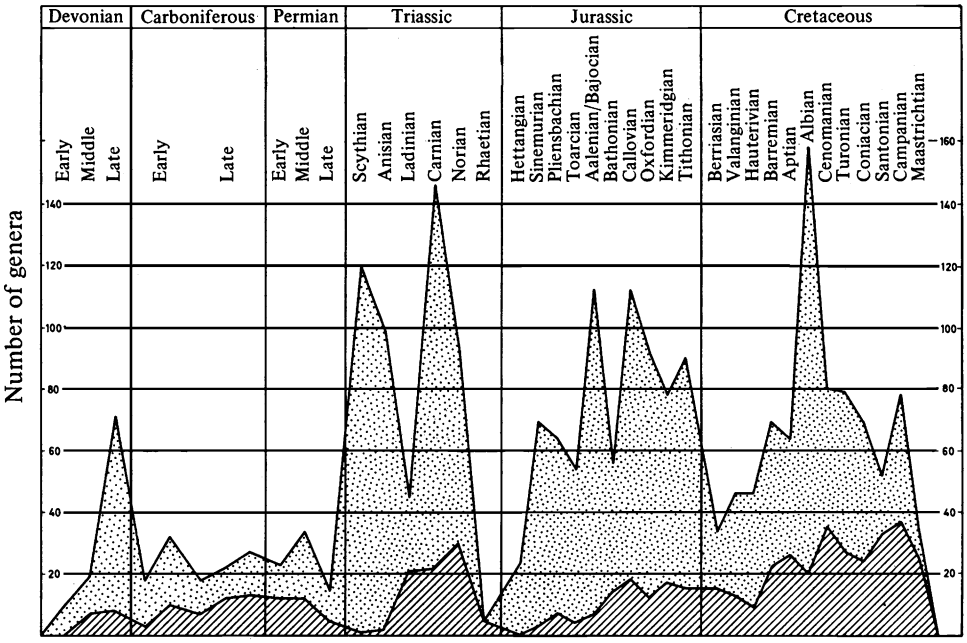


FIGURE 1 The differentiation of the ammonoids from the Devonian to the Cretaceous, based on the number of genera. Strippled area, new genera; Hatched area, continuous genera. From Lehmann, 1981, reprinted with the permission of Cambridge University Press.

been attributed partly to the extinction of many reptilian groups at the end of the Cretaceous (Simpson, 1953). The parallel adaptive radiation of marsupials in Gondwana has also been attributed to the Cretaceous extinctions and subsequent opening of ecological space (Springer *et al.*, 1997). However, within each lineage (placentals and marsupials) key innovations may have been involved: The radiation of ungulates and ruminants is associated with the opening up of the savannas (Fig. 2) but would not have happened if the organisms did not develop the morphological and physiological features necessary to exploit the habitat. Similarly, the radiation of the diprotodontians appears to have commenced in the Eocene and may have been promoted by a key adaptation for herbivory (Springer *et al.*, 1997). The actual basis for radiations subsequent to extinction episodes is still a subject of debate, particularly because coincidence between extinction events and subsequent radiations is generally poor. Vermeij (1995) argued that there is a stronger coincidence of species diversification episodes with increases in nutrient input into the biosphere.

We consider factors underlying species proliferation under two headings: intrinsic factors, and the concept of "key innovations," and extrinsic factors, including environmental change and colonization of isolated landmasses.

A. Intrinsic Factors: Key Innovations

Simpson (1953) suggested that the evolution of a suite of traits, or key innovations, that increase the efficiency with which a resource is used might allow species to enter a "new" adaptive zone, and the ecological opportunity thus allowed might promote diversification. The concept of the key innovation is an essential element in hypotheses of the evolution of specialization and subsequent adaptive radiation in herbivorous insects. However, the nature of key innovations is not often clear. In an attempt to define more clearly the concept, Berenbaum *et al.* (1996) examined cytochrome P450S and its relation to the adaptive radiation of butterflies. They found high levels of diversification in substrate recognition sites between species that do not share the same set of host plants; the reverse was true for those species that do share host plants. This result was taken to indicate that specialization may necessitate conservation of this region of the genome and could therefore be considered a key innovation.

Many attributes of species have been proposed as key innovations, or characteristics that have allowed diversification and proliferation. They generally involve the development of features that modify biotic interactions. Particular examples include the development of toxicity in plants (that allows them to "escape" preda-

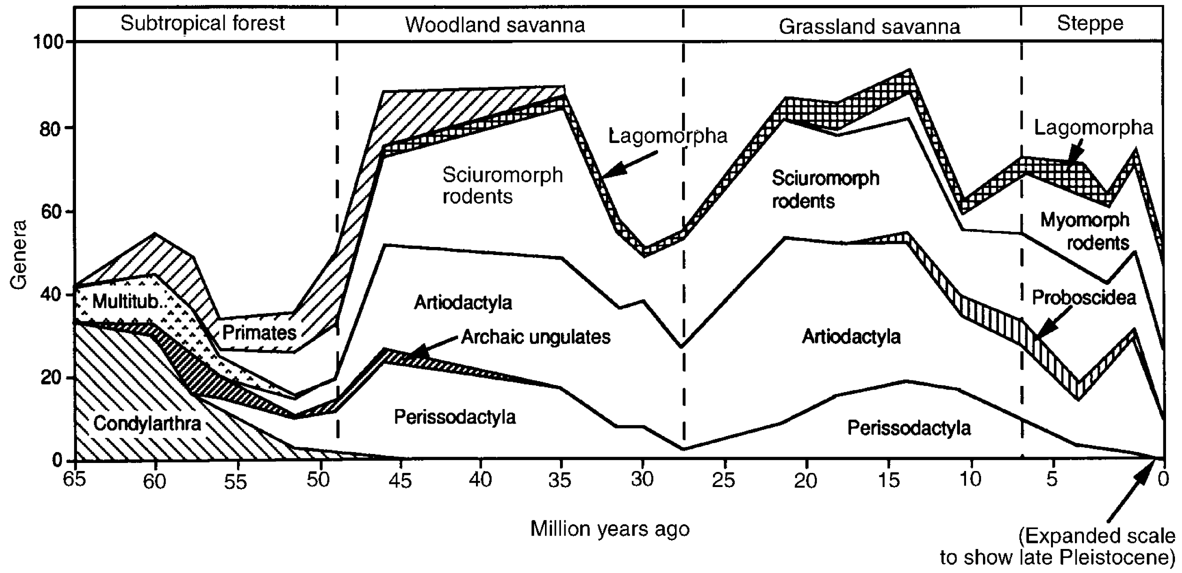


FIGURE 2 Plot of herbivore primary consumer diversity, taxonomic composition, and inferred dominant habitat type in North America for successive land-mammal ages during the Cenozoic. Vertical dashed lines indicate boundaries. From MacFadden, 1992. Fossil horses: systematics, paleobiology, and evolution of the family Equidae. Reprinted with the permission of Cambridge University Press.

tory pressure of insects) and subsequent development of tolerance to the toxin in insects which allows them to radiate onto the plants. Symbioses are another possible "evolutionary innovation," allowing the abrupt appearance of evolutionary novelty (Margulis and Fester, 1991). They provide a possible avenue through which taxonomic partners can enter into a new set of habitats unavailable to one or both of the symbiotic partners alone. One of the most famous examples is the radiation of ruminants in the African savannas, which has been attributed partially to the development of gut endosymbionts and the concomitant ability to digest cellulose. Among the Foraminifera, Norris (1996) showed that photosymbiosis appeared in the fossil record in synchrony with the taxonomic differentiation of three of the dominant surface water foraminifera groups in the Paleocene and early Eocene. This radiation was not paralleled in the asymbiotic sister group. Symbiosis was suggested to provide a jump-start for diversification by providing the ecological opportunity.

In their classic paper, Ehrlich and Raven (1964) examined how interacting species in themselves may create ecological opportunity, and hence periodically enhance evolutionary rates, through a broad "coevolutionary" response. They hypothesized that, when plant lineages are temporarily freed from herbivore pressure via the origin of novel defenses, they enter a new adaptive zone in which they can undergo evolutionary radiation. However, if a mutation arose in a group of insects that allowed it to feed on one of these previously protected lineages of plants, it would also be free to diversify in the absence of competition. Ehrlich and Raven envisioned this as a step-like process in which the major radiations of herbivorous insects and plants have arisen as a consequence of repeated opening of novel adaptive zones that each has presented to the other over evolutionary history. This idea, termed the "escalation/diversification" hypothesis (Berenbaum and Feeny, 1981), has been supported by the work of Farrell and colleagues, who have studied insect diversification in the context of host plants (Fig. 3). Repeated evolution of angiosperm feeding in phytophagous beetles is associated with an increased rate of diversification (Farrell, 1998). Similarly, there is consistently greater diversity among plants in which latex or resin canals have evolved as protection against insect attack (Farrell *et al.*, 1991).

Since Ehrlich and Raven, there has been a tremendous amount of research on the role of coevolution in dictating patterns of diversification. One of the major avenues that this research has taken is the study of the extent to which the phylogenetic order of divergence among herbivores or parasites corresponds to that among their hosts as a result of "parallel diversification" (Farrel and Mitter, 1994). A strongly corresponding evolutionary history might suggest a coevolutionary re-

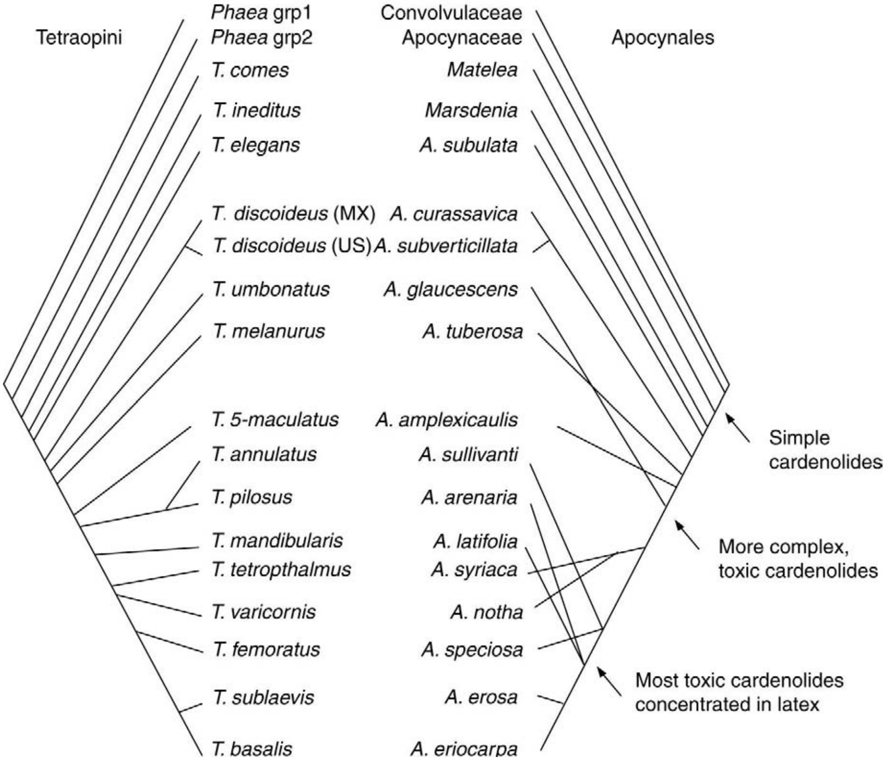


FIGURE 3 Phylogeny estimate of *Tetraopes* beetles based on morphology and allozymes, compared to literature based relationships of host plants. Host *Asclepias* shows an apparent progression toward increased complexity and toxicity of cardenolides, perhaps representing escape and radiation. From Farrell and Mitter, 1994.

sponse between the host and the herbivore or parasite. However, there appear to be few cases, at least among insect–plant interactions, in which the phylogeny agreement is precise. In most cases there has been periodic transfer of species to more distantly related hosts.

How do the coevolution arguments invoke adaptive radiation? The situation that Ehrlich and Raven (1964) envisioned was one in which the host radiated prior to exploitation and subsequent radiation by the herbivore ("escape and radiation"). This might be considered analogous to the opening up of an array of ecological opportunities every time the innovation arose for either "escape" or "exploitation." The established diversity of hosts could provide the necessary diversity of ecological roles and associated adaptations. Where coevolution involves parallel diversification, adaptive radiation may not be involved. In particular, parallel diversification might be considered analogous to geographic separation, with divergence of the host causing isolation of the herbivore. This might then be considered a nonadaptive radiation. On the other hand, parallel diversification might cause an escalation in responses, with enhanced toxicity and reciprocal tolerance evolving in step-like progression (Berenbaum, 1983; Farrell and Mitter, 1994). In this latter scenario, adaptive radiation can be implicated for both the herbivore and the host.

B. Extrinsic Factors

Speciation rates are generally considerably higher in novel environments, whether a lake in the middle of a continent or an island in the middle of the ocean (Schluter, 1998; Fig. 4).

1. Environmental Change

Environmental change has frequently been implicated in species radiations, with the opening up of new habitat. Diversification has frequently been suggested to occur under stressful conditions [e.g., for the origin of angiosperms (Shields, 1993) and the recent diversification of mole rats (Nevo *et al.*, 1984)]. However, any novel environment in which the organism is subjected to a new selective regime could be considered "stressful." In other words, an organism that successfully in-

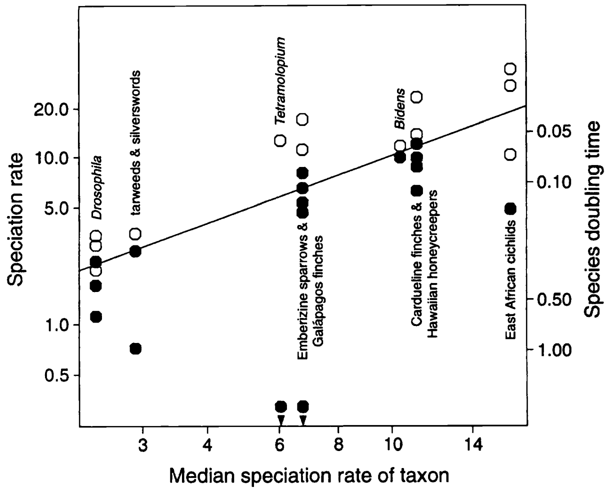


FIGURE 4 Per capita speciation rates of clades in novel environments (○) and in closely related "control" lineages inhabiting other environments (●). Rate estimates (y, left axis) are plotted against a dummy variable, the median of y-values. The solid line indicates $y = x$; points above the line therefore exhibit high rates. Rates were calculated from phylogenies based on allozyme frequencies. Time is measured in units of genetic distance (D). The calculation of rate y assumes exponential growth of species number: $y = \ln(N)/t$, where N is the number of extant species in a clade and t is its estimated time of origin. Corresponding times required for species number to double are indicated on the right; the number of species in a clade doubles after ln (2)/y time units. From *Endess Forms: Species and Speciation*, ed. by D. J. Howard and S. H. Barlocher, © 1998 by Oxford University Press, Inc. Used by permission of Oxford University Press, Inc.

vades a novel environment will inevitably be subject to stressful conditions. However, no matter whether the novel environment is considered stressful or simply a situation in which the organism is subject to a novel set of selective forces, it does appear to be associated with acceleration in evolutionary rates (Nevo *et al.*, 1984; Shields, 1993).

The evolution and adaptive radiation of the African cichlids (Fig. 5) appear to have been initiated by environmental change. Geological activity 20 million years ago (mya) caused the rivers in the area to become progressively meandric and swampy while still connected to the Zaire hydrological system. Over time, a mosaic of small, shallow, and isolated lakes developed, and finally the drainage system became closed and the lakes deepened (approximately 5 mya). The diversification of cichlid fish appears to have been initiated when river species moved into the swamp (Sturmbauer, 1998), and then successive radiations were associated with the development of protolakes and subsequently deep lakes.

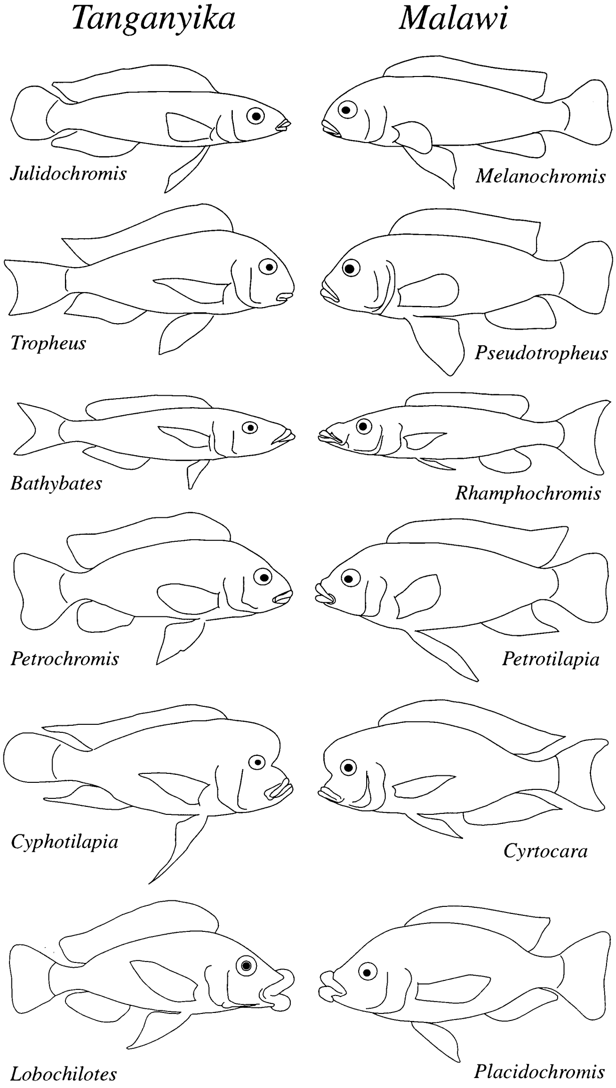


FIGURE 5 Adaptive radiation of cichlid fish, showing convergence in body shapes in Malawi and Tanganyika species. From TREE review, Meyer, 1993.

In geological history, environmental changes appear to form the basis of the Phanerozoic revolutions (Vermeij, 1995): Increasing temperature and nutrient supplies as a result of submarine vulcanism may have triggered later Mesozoic and perhaps early Paleozoic diversification episodes. Similar factors may underlie the iterative radiations of ammonoids throughout the geological record (Dommergues *et al.*, 1996). Each radiation appears to have originated from a few taxa, which went on to produce a wealth of morphological diversity. Although well documented for ammonoids, the pattern of iterative radiation is extremely rare in the fossil record. Within the total morphospace defined for the first three Jurassic stages, the radiation corresponds to a string of "events" separated by episodes of morphospace collapse. During each event, only a portion of morphospace (<45%) was filled; some morphs were reiterated, but each event had its own particular derived morphs.

2. Colonization of Isolated Landmasses (Particularly Oceanic Islands)

Isolated landmasses that have never been in contact with a source provide abundant opportunity in terms of newly available "niche space" for those taxa that manage to colonize them. Isolated archipelagoes are generally considered in a different category of adaptive radiations, although they are really just special cases of environmental change: the appearance of a new environment, which happens to be isolated in the ocean. This has much in common with the formation of, for example, a lake in the middle of a continent. In either case, newly created habitats that are isolated from a source of colonists provide an extraordinary opportunity for adaptive radiation. Both the novelty and the isolation are key features in allowing adaptive radiation in such areas. If a new habitat appears in close proximity to other such habitats, it will be colonized by taxa from those habitats. Species diversity patterns will then match closely the predictions of the MacArthur–Wilson model of island biogeography (MacArthur and Wilson, 1967); that is, species diversity patterns will be governed by ecological processes. As isolation from the source of colonists increases, fewer taxa will be able to colonize the new habitat, and the low rate of colonization may provide sufficient time for species diversification to occur. Adaptive radiations are most likely to occur at the extreme ends of the dispersal range of a given taxon (Whittaker, 1998).

IV. ARE CERTAIN TAXA MORE LIKELY TO UNDERGO ADAPTIVE RADIATION THAN OTHERS?

Are species predisposed to undergo adaptive radiation because of a broad environmental tolerance, generalized feeding patterns, or perhaps some proclivity to develop novel associations? This question has been developed by some authors. For example, Adler and Dudley (1994) compared patterns of adaptive radiation among birds and butterflies in the insular Pacific: Birds have undergone extensive adaptive radiation, whereas butterflies have not. They argued that speciation in butterflies may be constrained by the mechanics of insect–plant coevolution that prevents rapid diversification. However, this argument is not well supported because other insects with similar coevolutionary ties have undergone some of the most spectacular insular adaptive radiations known. It appears that almost any group of organisms is capable of undergoing adaptive radiation upon being provided ecological opportunity that it can exploit.

V. HOW DOES ADAPTIVE RADIATION GET STARTED?

A. Initiation of Adaptive Radiation: Genetic Changes

1. Founder Events

The establishment of species in new environments inevitably involves sampling from the parent population. The size of the sample that can build a new population can be very small (cf. founder effects), although it need not necessarily be so. In particular, if, subsequent to colonization, a very small number of individuals were to proliferate rapidly, there would be little subsequent loss in genetic variability (Nei *et al.*, 1975). Consequently, the deleterious effects of inbreeding are largely mitigated. However, because the genes represented in the founding population are only a small sample of the original population, genetic drift may be pronounced.

The nature of genetic changes during shifts in population size, particularly those experienced during or after population bottlenecks, has been the subject of considerable controversy in recent years. Clearly, a crash in population size as a result of a genetic bottleneck or founder event will cause allele frequencies at some loci to differ from those of the parent population because of accidents of sampling (Templeton, 1980). The debate concerns the nature of genetic changes that occur subsequent to the bottleneck, during the period of population growth. Traditional arguments suggested that founder events may trigger rapid species formation (Carson and Templeton, 1984). However, recent arguments have largely refuted the contribution of founder events to reproductive isolation (Barton, 1996).

Other possible changes during founder events are due to genetic reorganization. Carson (1990) proposed that blocks of loci are destabilized when a newly founded colony undergoes a flush of exponential growth, during which time selection is relaxed and recombinants that ordinarily have low fitness survive.

Release of additive genetic variance through change in epistatic interaction may allow formation of novel recombinants (Goodnight, 1988). Recent work has questioned the importance of epistatic interactions and the nature of "destabilization" among blocks of loci during periods of reduced population size. Slatkin (1996) discussed changes that might occur subsequent to reduction in population size in terms of conventional population genetics. Genetic drift will be the primary force causing genetic changes during the bottleneck. However, when the population starts to grow, the influence of genetic drift is expected to become weaker so that selection is most effective during and immediately after the period of rapid demographic expansion. On the other hand, Otto and Whitlock (1997) argued that, although it is true that the probability of fixation of beneficial alleles present during the bottleneck is increased during the subsequent period of population growth, a large proportion of alleles are lost during the bottleneck and few new mutations can occur while the population is at small size. The resultant effect of these opposing forces is that the number of beneficial mutations fixed per generation remains virtually unchanged by the bottleneck. Nevertheless, selection subsequent to a genetic bottleneck has an important effect on alleles that are initially rare and that would tend to be lost to stochastic events in populations of constant size (Slatkin, 1996). In addition, there is evidence that stress on a genetic system (e.g., as a result of a population crash) may activate transposable elements (Carson, 1990), which can exercise a mutagenic effect by interrupting structural or regulatory regions of the genes into which they become integrated. Therefore, mutation rates may actually be higher in founder populations, providing raw material on which selection can act, and which in turn could lead to a rapid recovery of genetic variability. There are thus two processes which can be associated with colonization events: (i) possible genetic changes/restructuring within the population and an expected loss of genetic variability and (ii) the subsequent recapturing of genetic variation during population expansion as a result of selection (Slatkin, 1996).

2. Rapid Proliferation and Hybridization

Differential mixing of characters during segregation of populations and species may occur as a result of hybridization of newly divergent taxa (Harrison, 1993). Behavioral changes during founder events may facilitate hybridization because it has been suggested that sexual interactions may lose specificity subsequent to a founder event (Kaneshiro, 1989). Closely related heterospecifics may therefore hybridize and/or introgress subsequent to colonizing a new landmass, but the extent of genetic exchange may differ between regions of the genome (DeSalle and Giddings, 1986). Differences may be particularly pronounced between character sets involving nuclear and extranuclear DNA, such as mitochondrial or chloroplast DNA. Extranuclear DNA differs from nuclear DNA because of its greater sensitivity to the effects of each founder event due to its smaller effective population size (one-fourth) relative to nuclear DNA, attributable to its transmission primarily through the female line, and its existence as a single copy (Avise, 1991). Analysis of extranuclear DNA information will result in a gene genealogy, but it is likely to provide an incomplete history of the organisms if much hybridization has taken place. The occurrence of hybridization may explain differences that have been found between nuclear and mitochondrial phylogenies, particularly for the Hawaiian *Drosophila* (DeSalle *et al.*, 1997). Indeed, natural hybridization with the formation of fertile hybrids has been documented between closely related species of *Drosophila* on the youngest island of the Hawaiian Islands (Carson, 1989). Among silverswords, hybridization has been implicated as an important element in the adaptive radiation of the group (Baldwin, 1997).

Among the Galapagos finches, recent molecular data have failed to distinguish species limits, at least for the morphologically defined ground and tree finch species: Individuals representing different morphologically identified species are intermingled. This may be explained by interspecific hybridization and/or sorting of haplotypes (Freeland and Boag, 1999). In the case of ground and tree finches, both explanations may apply. The incomplete species differentiation within these taxa may be taken as an indication that adaptive radiation is currently ongoing.

Failure of molecular data to distinguish species limits has been found in several other adaptive radiations. For the cichlid fish of Lake Malawi, it has been suggested that speciation is occurring faster than alleles can become fixed within a species (Moran and Kornfield, 1993). Among Hawaiian crickets, Shaw (1996) comments on a discrepancy between phylogenies generated on the basis of song (Otte, 1994) with that generated from mtDNA variation. Although the discrepancy could be explained by problems with current taxonomic boundaries, and problems with mtDNA lineage sorting, she argues that hybridization and introgression appear to be the most likely explanation.

Although the number of species in which hybridization has been documented is currently small, it is likely to increase as researchers accumulate phylogeographic knowledge based on multilocus molecular data for different radiations.

B. Initiation of Adaptive Radiation: Ecological Changes

Populations frequently initiate a cycle of change in abundance and distribution upon colonization of a novel habitat or an unoccupied set of niches, during which they undergo ecological release, expand their range, and adopt a more generalized habit (Cox and Ricklefs, 1977). After their initial rise to dominance as widespread generalists, these taxa may subsequently be competitively displaced from much of their original range by younger relatives (Darlington, 1957). Wilson's (1961) "taxon cycle" was used to describe such changes in the distributional pattern of Melanesian ants. He suggested that widespread, dispersive populations give rise to many more restricted and specialized species. Regular changes in ecological and geographical distribution have been recognized in many island systems (Cox and Ricklefs, 1977).

Ecological release may be the precursor to adaptive radiation. This argument has been supported by recent experimental evidence: Rainey and Travisano (1998) studied adaptive radiation experimentally using the bacterium *Pseudomonas flourescens.* This interesting bacterium is known to evolve rapidly in novel environments, with evolutionary differences detectable in the morphology and ecological affinity of the phenotype. An isogenic population of one morph was propagated in (i) a spatially heterogeneous environment and (ii) a spatially homogeneous environment. In the spatially heterogeneous environment, extensive morphological diversification was found to occur within 3–10 days, resulting in three dominant morphs. This was taken as evidence for diversifying selection. In the spatially homogeneous environment, no morphological variation was found.

VI. THE PROCESSES OF ADAPTIVE RADIATION: CASE STUDIES

What are the factors underlying diversification in an adaptive radiation? This question has fascinated biologists for many years. However, until recently, the characters involved in the adaptive radiation necessarily had to be used as the basis for phylogenetic inference. Given that adaptive radiations are characterized by tremendous levels of convergence in almost every morphological and ecological character, it has been very difficult to interpret evolutionary processes. As Givnish (1997) notes, "any rigorous, noncircular study of adaptive radiation must be based on a phylogeny that has been derived independently of the traits involved in that radiation." Recent advances in molecular techniques, although not without problems of their own, provide an opportunity to obtain such an independent assessment of evolutionary history (see review of arthropod radiations in Hawaii in Roderick and Gillespie, 1998). The following sections outline case studies from a range of taxa, most of which have used a combination of morphological and molecular techniques in an attempt to gain some understanding of processes underlying adaptive radiations.

A. Galapagos Finches

There are currently 14 recognized species of Darwin's finches in six genera, which have evolved from a common ancestor (Fig. 6; Lack, 1947; Grant, 1986). Of these, 13 live in the Galapagos Islands. Based on morphological, behavioral, and ecological data, they have been divided into three lineages: First, the ground finches, *Geospiza* (6 species), which are found in more arid areas of the archipelago and feed on seeds on the ground. Three of the species which are considered the most "finch-like" differ primarily in body and beak size and are known as the large (*G. magnirostris*), medium (*G. fortis*), and small (*G. fuliginosa*) ground finches. The 3 other species of ground finches have longer beaks. Two feed on cactus flowers and pulp as well as seeds and are known as the large (*G. conirostris*) and small (*G. scandens*) cactus ground finches. Finally, the sharp-beaked ground finch (*G. difficilis*) supplements its diet with the eggs and blood of other birds and reptile ticks. Second, the tree finches, which are found mostly in trees and shrubs, are divided into 3 genera: *Cactospiza* [the woodpecker finch (*C. pallida*) and the mangrove finch (*C. heliobates*)], *Camarhynchus* [the large tree finch (*C. psittacula*), the medium tree finch (*C. pauper*), and the small tree finch (*C. parvulus*)], and *Platyspiza* (the vegetarian finch, *P. crassirostris*). All except *P. crassirostris* are insect eaters. Finally, the warbler-like finches, which are small with slender beaks, are in 2 genera: The warbler finch (*Certhidea olivacea*) catches insects like a warbler, and the Cocos finch (*Pinaroloxias inornata*) is the only Darwin finch that lives outside the Galápagos Archipelago. It appears to have colonized Cocos Island from the Galapagos.

Darwin's finches share common features of nest architecture, egg pattern, and courtship displays. They

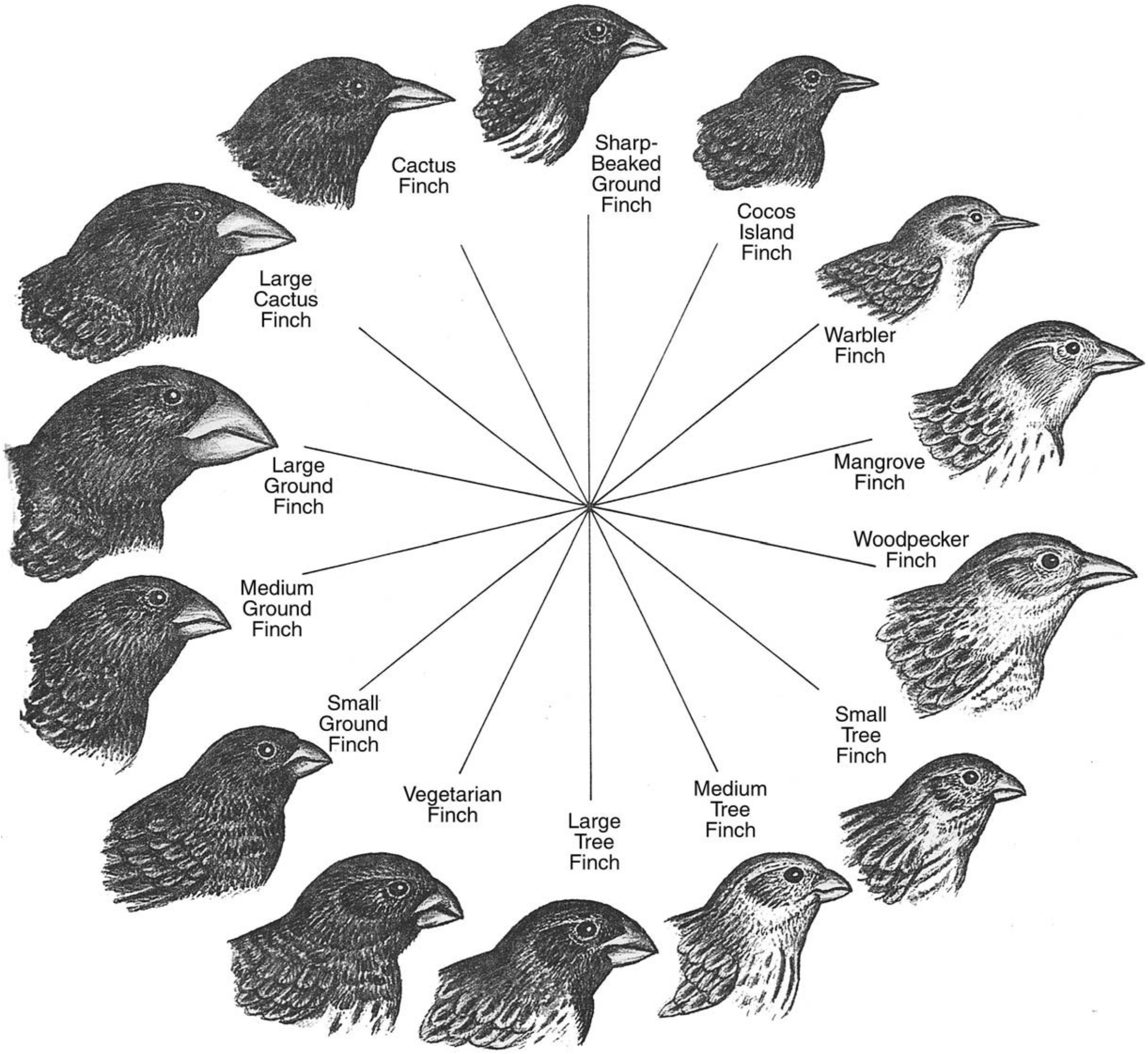


FIGURE 6 Adaptive radiation of Darwin's finches. From Grant, 1991.

differ in song, morphology, and plumage. Based on morphology, allozyme, and DNA sequence data, the warbler finch *C. olivacea* appears to be closest to the ancestral form. However, recent molecular data indicate that the Cocos finch *P. inornata* is closest to the tree finches of the Galapagos (Sato *et al.*, 1999; Petren *et al.*, 1999). Also, the vegetarian finch appears to be ancestral to both the tree and ground finches rather than being a member of only the tree finch group.

Grant (1986) proposed a three-step process to speciation to explain the adaptive radiation of the group: (i) The ancestral species arrives in the archipelago; (ii) the species spreads to other islands and as a result of this, there will probably be some selection and hence differentiation in size because the islands are different; and (iii) members of the original and derived populations encounter each other, and as a result of competition for food and selection against intermediates character displacement causes rapid divergence in feeding structures between the species when they come together. Therefore, the radiation appears to be based on (i) the isolation of the archipelago, which provided ecological opportunity; (ii) considerable distances between the different islands, which has led to infrequent interisland exchange; and (iii) different environments on the different islands, which have selected for different feeding niches both within and between islands.

B. Hawaiian *Drosophila*

The Hawaiian Drosophilidae represent some of the most striking examples of adaptive radiation known for any group anywhere (Kaneshiro, 1988), with 337 species in the genus *Drosophila* (pomace flies), 19 in *Idiomyia* (picture-winged flies), 122 in *Scaptomyza,* and 11 in *Titanochaeta* (Hardy, 1965; Hardy and Kaneshiro, 1981). Based on the premise that founder events are the most important mechanism of speciation in the Hawaiian *Drosophila,* Kaneshiro (1983) proposed models suggesting the importance of sexual selection in driving species proliferation in these insects. The Hawaiian Drosophilidae, particularly the males of the picture-winged species, often have ornately patterned wings as well as unusual modifications of the mouthparts and legs. The extraordinary manifestations of these features in the male picture-winged species are frequently accompanied by elaborate courtship behavior. In classic studies of picture-winged *Drosophila,* Kaneshiro (1983) found evidence for asymmetrical sexual isolation: Females from a geologically older island were found to be highly discriminating in terms of mate choice and would not mate with a closely related species from a younger island. In contrast, females from species on the younger island readily accepted males from the older island species. Based on these observations, Kaneshiro proposed that courtship requirements are relaxed during the early stages of colonization of a new island; sexual behavior may then become less constrained and simpler, and there may be more intraspecific variability. In such circumstances, there would be strong selection for less discriminating females because of the difficulties in finding mates and reproducing when the population size is small. Intrasexual selection may then operate to cause divergence of the sibling species during isolation because of a shift in the distribution of mating preferences during the founder/flush cycle (Carson, 1986). The result would be a shift in the mating system, which could then be fixed at a new "equilibrium."

Evolution of the Hawaiian *Drosophila* is commonly treated as an example of sexual selection influencing speciation (Carson, 1986). The nature of this influence is not entirely clear. *Drosophila heteroneura* and its close relative, *D. silvestris,* are partially sympatric in forests on the island of Hawaii, where they occasionally hybridize. The species are distinguished by the much broader head of *D. heteroneura*. Male aggression and male courtship are the major determinants of male mating success. Accordingly, it appears that male head width is subject to sexual selection through mate choice, although this, and perhaps related sexually dimorphic traits, may not be involved in behavioral isolation (Boake *et al.*, 1997).

A recent molecular phylogeny of picture-winged Hawaiian *Drosophila* showed that clades are characterized according to whether they breed on fungi, leaves, fruit, or bark (Kambysellis and Craddock, 1997). Suites of reproductive characters (in particular, ovarian egg and ovipositor traits) appear to have evolved together (Kambysellis, 1993), and adaptive shifts to new breeding sites appear to have been important at least in the early diversification of the group and in some recent speciation events (Kambysellis and Craddock, 1997). The interplay of such ecological shifts with sexual selection in allowing diversification of Hawaiian *Drosophila* remains unresolved.

C. Hawaiian Honeycreepers

The endemic Hawaiian subfamily Drepanidinae, or Hawaiian honeycreepers, includes 33 historically known species and at least 22 "fossil" species (Fig.

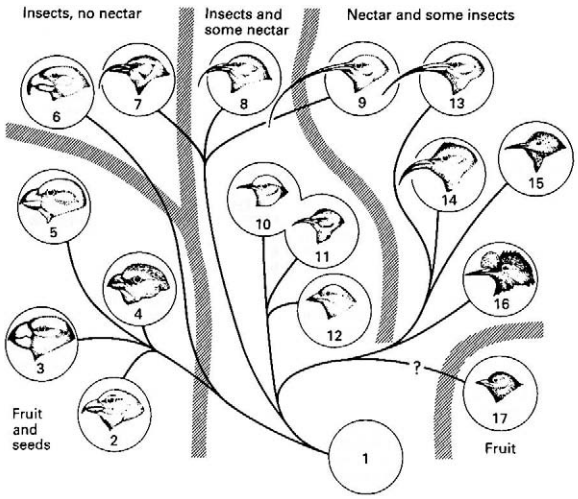


FIGURE 7 The inferred pattern of evolution of dietary adaptations as represented by 16 of the Hawaiian honeycreepers. 1. Unknown finch-like colonist from Asia; 2. *Psittirostra psittacea*, 'O' u; 3. *Chloridops kona*, Kona Grosbeak (extinct); 4. *Loxioides bailleui*, Palila; *Telespyza cantans*, Laysan Finch; 6. *Pseudonester xanthophyrs*, Maui Parrotbill; 7. *Hemignathus munroi*, 'Akiapola'au; 8. *H. lucidus*, Nuku Pu'u; 9. *H. obscurus* 'Akialoa; 10. *H. parvus*, 'Anianiau; 11. *H. virens*, 'Amakihi; 12. *Loxops coccineus*, Hawaii 'Akepa; 13. *Drepanis pacifica*, Mamo (extinct); 14. *Vestiaria coccinea*, liwi; 15. *Himatione sanguinea*, 'Apapane; 16. *Palmeria dolei*, 'Akohekohe; 17. *Ciridops anna*, 'Ula-'aihawane (extinct). From Cox, Biogeography: An ecological and evolutionary approach (5th ed.) Blackwell Science, Ltd., with permission.

7; James and Olson, 1991). The three tribes are (i) the Psittirostrini, Hawaiian finches (23 species), which includes 5 extant species, 4 historically extinct species, and 14 fossil species; (ii) the Hemignathini, Hawaiian creepers and allies (22 species), which includes 13 extant species, 3 species and 2 subspecies historically extinct, and 6 fossils; and (iii) the Drepanidini, Mamo's, Iiwi's, Apapane's, crested honeycreeper and allies (9 species), which includes 4 extant species, 3 species historically extinct, and 2 fossils. Based on the tremendous morphological and ecological differentiation known from historical and recent collections, these birds have been heralded as one of the best known examples of adaptive radiation. However, the high proportion of currently extinct species makes it difficult to develop hypotheses regarding processes underlying the adaptive radiation. Molecular studies based on extant species suggest that differentiation has occurred between islands only (Tarr and Fleischer, 1995). One might speculate that species have proliferated in a manner similar to that proposed for the finches in the Galapagos, as described previously.

D. African Cichlids

Lacustrine fish represent some of the most spectacular cases of adaptive radiation in vertebrates (McCune, 1997), with the best known being cichlids. Cichlids are spiny-rayed freshwater fishes that reach their most abundant diversity in Africa, particularly in the great east African lakes of Victoria (>400 species), Malawi (300–500 species), and Tanganyika (approximately 200 species) (Fig. 5). In each of these lakes the fish exhibit spectacular diversity in trophic morphology, including specialist algal scrapers, planktivores, insectivores, piscivores, paedophages, snail crushers, and fin biters. In addition to their trophic diversity, they exhibit a striking array of color patterns. They also show complex mating behaviors, polyandrous mating systems, and a tendency to breed in leks. Several factors are considered to be involved in the diversification of these fish. One suggests that morphological adaptation is the primary event underlying speciation (Liem, 1973). The basis for this assertion is their trophic apparatus, which is unique in the possession of two sets of jaws—one in the mouth for sucking or scraping and the other in the throat for macerating, crushing, or piercing. These structures can be modified according to the diet of the fish. The second factor considered to be involved in allowing reproductive isolation and hence diversification of the fish is sexual selection, which allows subsequent morphological differentiation.

Recent molecular work (Meyer, 1993; Meyer *et al.*, 1996) has shown that cichlids in any one lake are far closer to each other than to morphologically similar species in other lakes. Evolution has generated similar ecomorphs in each of the lakes. There has been considerable debate regarding the nature of the mechanism underlying the proliferation of species within a lake, particularly whether it is considered "allopatric," "microallopatric," or "sympatric" (McCune and Lovejoy, 1998). However, perhaps the most broadly accepted mechanism involves repeated isolation. For example, Lake Tanganyika has apparently undergone repeated decreases in water level (as much as 2000 ft). Populations may have been isolated in small pockets of water for long enough to become reproductively isolated. Sexual selection may play a role in allowing reproductive isolation of the fish subsequent to isolation because closely related species differ primarily in color pattern and not in jaw morphology (Albertson *et al.*, 1999). It appears that morphological adaptation played a role in the early radiation of the group, resulting in distinct clades that differ in jaw morphology; only minor morphological differences are observed within genera (Stauffer *et al.*, 1997). Recent diversification is largely associated with variation in adult male color pattern.

E. Hawaiian Silverswords

The Hawaiian silversword alliance has been considered "the best example of adaptive radiation in plants" (Raven *et al.*, 1992). It consists of 30 species in three genera (*Wilkesia, Dubautia,* and *Argyroxiphium*) but with one common ancestor. Life form diversity includes trees, shrubs, mat plants, monocarpic and polycarpic rosette plants, cushion plants, and vines that occur across a broad environmental spectrum from rain forests to desert-like settings (Carr, 1985). Major ecological shifts have accompanied speciation in each of the major island-endemic lineages (Fig. 8; Baldwin, 1997). The estimated minimum rate of diversification in the silversword alliance (0.56 ± 0.17 species per million years) is comparable to, or higher than, rates of several more ancient continental groups that have been regarded as examples of adaptive radiation (Baldwin and Sanderson, 1998).

F. Lizards in the Caribbean

Currently, 139 species of Caribbean lizards are recognized, 80% of which occur in the islands of the Greater Antilles (Cuba, Hispaniola, Jamaica, and Puerto Rico)

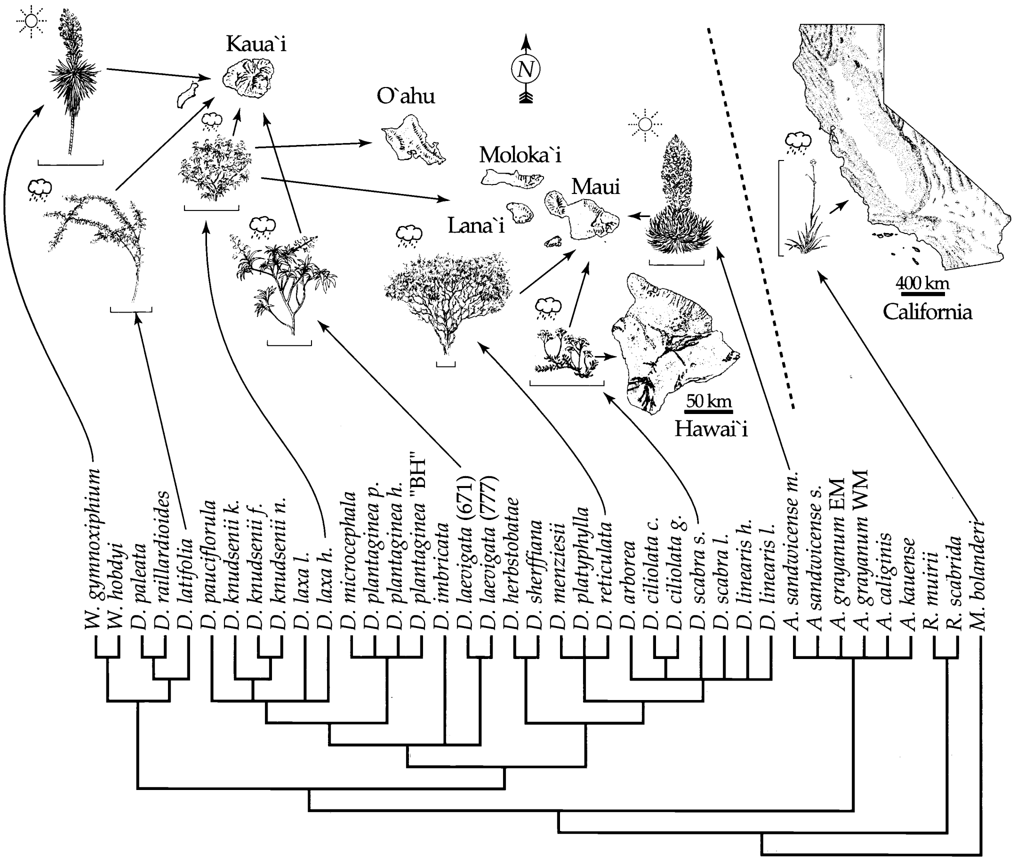
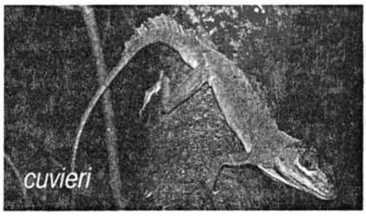


FIGURE 8 Phylogeny of the Hawaiian silversword alliance and the American tarweed genera *Madia* and *Raillardiopsis* based on DNA sequences. Generic abbreviations: A, *Argyroxphium*; D, *Dubautia*; M, *Madia*; R, *Raillardiopsis*; W, *Wilkesia*. EM, East Maui; WM, West Maui. Plant scale bars equal 1 meter. Arrows indicate island(s) or State (California) of occurrence, not necessarily the location within islands or California where the plants are found. Plant illustrations depict a variety of habits and habitat/island occurrences in the silversword alliance. Sun symbol, restriction to dry habitats; rain-cloud symbol, restriction to wet habitats (including bogs). From Baldwin, 1997, reprinted with the permission of Cambridge University Press.

(Fig. 9). Losos (1992) and Losos *et al.* (1998) found that on each of the islands lineages have diversified in such a way as to occupy a range of ecological roles, with as many as 11 species occurring sympatrically. Different species live, for example, on twigs, in the grass, or on tree trunks near the ground. Twig anoles tend to be slender with short legs and tails, whereas the trunk-ground anoles are stocky with long legs and poorly developed toepads. Losos and colleagues recognized six habitat specialists and called these "ecomorphs," which are recognizable on the basis of morphological measurements. The same set of ecomorphs are found on each island (with a few exceptions): Four occur on all islands (trunk-ground, trunk-crown, crown-giant, and twig), one occurs on all islands except Jamaica (grass-bush), and one occurs only on Cuba and Hispaniola (trunk). Phylogenetic studies using mitochondrial DNA indicate that the same ecomorphs have evolved independently on each island, i.e., they appear to have arisen as a result of one-to-one convergence of the same set of ecomorph types on each island (Losos, 1992). However, the sequence by which they evolved differs on each island. Such tests of convergence are among the strongest available for evaluating the premise that the number and types of coexisting species are locally determined. The conclusion from this research was that "adaptive radiation in similar environments can overcome historical contingencies to produce strikingly similar evolutionary outcomes" (p. 2115, Losos *et al.*, 1998).

Tree Crown
Large body, large toe pads

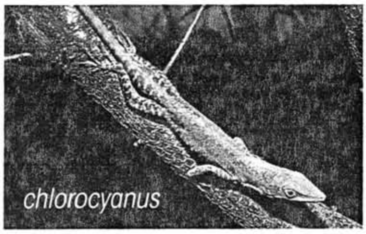
Cuba--*Anolis equestris*
Hispaniola--*A. ricordii*
Jamaica--*A. garmani*
Puerto Rica--*A. cuvieri*



Jonathan Losos

Upper Trunk/Canopy
Large toe pads, can change color

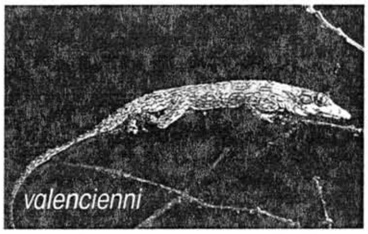
Cuba--*Anolis porcatus*
Hispaniola--*A. chlorocyanus*
Jamaica--*A. grahami*
Puerto Rica--*A. evermanni*



Kevin de Quieroz

Twig
Short body, slender legs &tail

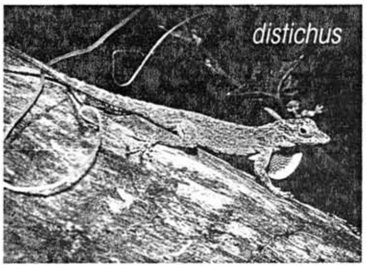
Cuba--*Anolis angusticeps*
Hispaniola--*A. insolitus*
Jamaica--*A. valencienni*
Puerto Rica--*A. occultus*



Jonathan Losos

Midtrunk
Long forelimbs, vertically flattened body

Cuba--*Anolis loysiana*
Hispaniola--*A. distichus*
Jamaica-- none found
Puerto Rica--none found



Kevin de Quieroz

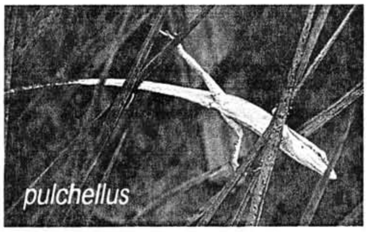
Lower Trunk/ Ground
Stocky body, long hind limbs

Cuba--*Anolis sagrei*
Hispaniola--*A. cybotes*
Jamaica--*A. lineatopus*
Puerto Rica--*A. gundlachi*

Robert Rattner

Grass/ Bush
Slender body, very long tail

Cuba--*Anolis alutaceus*
Hispaniola--*A. olssoni*
Jamaica--none found
Puerto Rica--*A. pulchellus*



Jonathan Losos

G. Sticklebacks in Deglaciated Lakes of Canada

Deglaciated lakes of coastal British Columbia, Canada, harbor a tremendous diversity in the form of undescribed sibling species of fish. There are numerous examples of species pairs in many fish families, and the repeated occurrence of such sympatric pairs has been attributed to novel ecological opportunity provided by deglaciation and recolonization from relatively depauperate faunas. In particular, the radiation of species of three-spined sticklebacks (a species complex currently classified under *Gasterosteus aculeatus*) includes some of the youngest species known—less than 13,000 years. No more than two species occur in any one lake, but pairs of species in different lakes appear to have evolved completely independently of other pairs. Results based on morphological, ecological, and molecular data suggest that species have diverged as a result of parallel bouts of selection for alternate trophic environments (Fig. 10; Taylor *et al.*, 1997). Natural selection has been implicated as the major cause of evolution (Schluter, 1994; Schluter and McPhail, 1993). This study has been very important in highlighting the role of divergent natural selection as a mechanism underlying adaptive radiation—a view held widely by naturalists in the earlier part of the century although, until recently, without much support (Schluter, 1996). Ecological character displacement appears to underlie diversification in the sticklebacks, with the evolution of reproductive isolation a by-product of resource-based divergent natural selection.

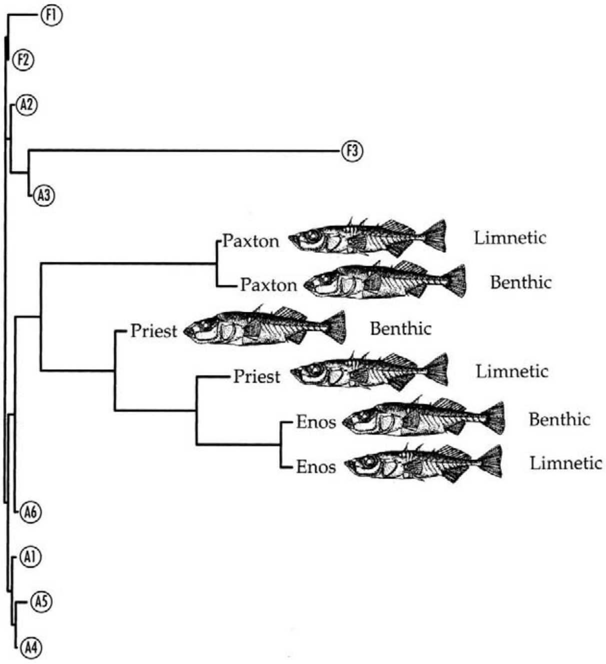


FIGURE 10 Ecological divergence in sticklebacks. A1–A6 are anadromous populations, F1 and F2 are populations resident in freshwater streams, and F3 is a solitary lake population. Benthic and limnetic species of fish are shown for Paxton, Priest, and Enos Lakes. Adapted from Taylor *et al.*, 1997, with the permission of Cambridge University Press.

H. Hawaiian *Tetragnatha* Spiders

The long-jawed, orb-weaving spider genus *Tetragnatha* comprises 295 described species worldwide. The adaptive radiation of the genus in Hawaii has been uncovered only recently, with a total of 28 species described and many more undescribed (Fig. 11; Gillespie *et al.*, 1998). This radiation spans a huge spectrum of colors, shapes, sizes, ecological affinities, and behaviors. Phylogenetic analyses have been performed on the "spiny-leg" clade of Hawaiian *Tetragnatha,* a lineage that has adopted a vagile, cursorial predatory strategy. On each island, species can be characterized as "green leaf-dwelling," "maroon," and "gray/black bark-dwelling." However, species on any one island are generally most closely related to each other, and each of the different ecomorphs appears to have evolved independently on the different islands (Gillespie *et al.*, 1997). There appears to have been a one-to-one convergence of the same set of ecomorph types on each island.

I. Hawaiian Swordtail Crickets

Thirty-five species of *Laupala* (Gryllidae) are found in rain forests across the Hawaiian archipelago, and up to 4 species coexist on any given island (Otte, 1994). In

FIGURE 9 Adaptive radiation of Carrbbean lizards. In the Greater Antilles, *Anolis* lizards that adapted to corresponding niches look alike, although they are not closely related. Shown is a sampler of niche holders listed by species name, with a photograph of one member of each category.

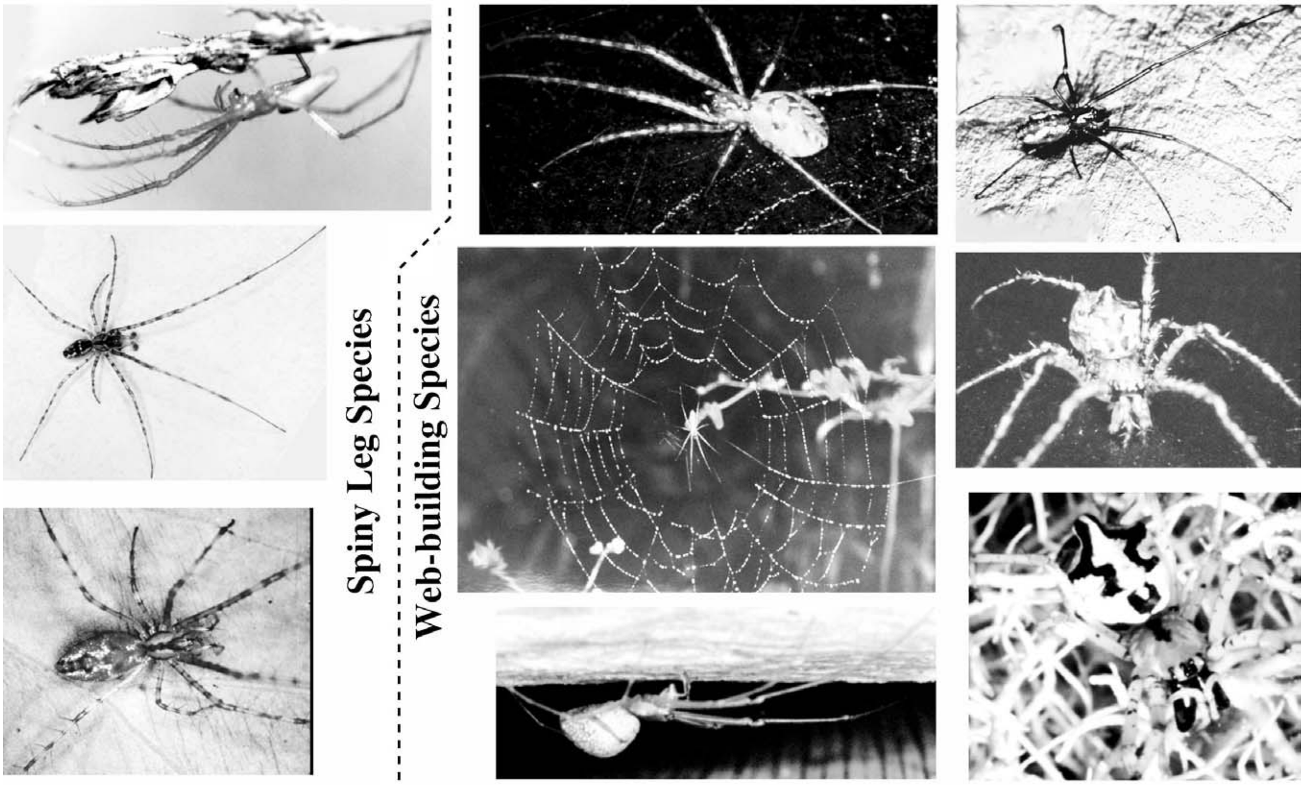


FIGURE 11 Adaptive radiation of Hawaiian *Tetragnatha* spiders. Shown are representative species from a lineage that has abandoned web building ("spiny-leg" species) and a large lineage of web-building species that show tremendous morphological and ecological diversity. Spiny-leg species: From top, *T. brevignatha* and *T. kamakou* from wet forests on E. Maui; *T. pilosa* from wet forests on Kauai. Web building species: Five from wet forests on E. Maui (clockwise from bottom left): *T. filiciphilia*, *T. stelarobusta*, *T. paludicola*, *T. eurychasma*, and *T. trituberculata*; (bottom right) undescribed species from wet forests on Hawaii.

common with other crickets, courting males "sing" to attract females by rubbing their forewings together. Each species has a unique song, and females respond preferentially and move toward the source of their species-specific song. It appears, therefore, that the song serves as an important mate recognition signal and variation in the acoustic system may serve as a basis for behavioral isolation between species. Phylogenetic analyses based on morphological (Otte, 1994) and molecular (Shaw, 1996) characters indicate that extensive intraisland species formation has occurred in this group. Differentiation between closely related species is associated with changes in song. These features of the *Laupala* system suggest that, unlike the Hawaiian *Drosophila,* founder events have not played a major role in the initial stages of species diversification (Shaw, 1996). Rather, differentiation appears to occur through the interaction of sexual selection on genetically well-structured populations. More work is needed to understand the relative importance of these factors in the initial divergence of species in this group.

J. *Partula* Land Snails

Land snails can exhibit tremendous phenotypic polymorphism with respect to the color, banding, and chirality of the shell. *Partula* is a group of species inhabiting the volcanic islands of the South Pacific. The evolution of this group parallels the geological history of the islands. They provide an example of classical adaptive radiation and also show that changes in ecologically important traits need not be accompanied by large genetic changes, and that the ecological changes can take place before reproductive isolation is complete (Johnson *et al.*, 1993; Murray *et al.*, 1993). Competition appears to be important in dictating the array of species at a site.

K. Bacteria in Culture

As described previously, experimental studies on a bacterium, *Pseudomonas flourescens,* have provided fascinating insights into possible processes underlying adaptive radiation. *Pseudomonas flourescens* was found to diversify rapidly from an isogenic population under conditions of environmental heterogeneity. Moreover, selection appeared to be the primary force maintaining diversity in the heterogeneous environment. The evolution of variant forms was found to follow a predictable sequence (cf. Caribbean lizards, stickleback fish, and *Tetragnatha* spiders), and competition is inferred to maintain the variation (Rainey and Travisano, 1998).

L. Conclusions

Three primary mechanisms have been inferred to underlie species proliferation in adaptive radiations: (i) geographical differentiation and subsequent character displacement (on the basis of either ecologically or sexually important characters) in Galapagos finches and African cichlids; (ii) sexual selection, with or without genetic bottlenecks, in Hawaiian *Drosophila* and *Laupala* crickets; and (iii) divergent natural selection based on ecological shifts in Hawaiian *Tetragnatha* spiders, plants in the Hawaiian silversword alliance, lizards in the Caribbean, *Partula* land snails in the Pacific, stickleback fish in deglaciated lakes, and bacteria in culture.

VII. THE FUTURE

Adaptive radiation, as a phenomenon, has tremendous research potential: The existence of a suite of closely related species adapted to exploit different habitats or lifestyles allows one to make comparative studies on the processes of speciation and selection in natural populations. Molecular systematics is providing a much better understanding of the evolutionary history of groups of closely related species and provides the opportunity for testing mechanisms underlying adaptive radiation (Givnish, 1997). The phylogenetic hypotheses provide a framework for examining the evolution of specific morphological, ecological, behavioral, and physiological adaptations and the circumstances in which they have arisen and have allowed adaptive radiation.

Adaptive radiations have recently become the focus of studies in conservation biology because they are frequently, particularly on islands, associated with high frequencies of endemism. As a corollary to the high endemism, many of the species that make up an adaptive radiation are often very rare and are characterized by very high extinction rates. Far more extinctions have been documented from islands than from continents: Of known extinctions, those from islands comprise 58% for mammals (most of which are absent from remote islands), 80% for mollusks, and 85% for birds (Whittaker, 1998). Similarly high extinction rates are known for radiations of lacustrine fish. The high rate of extinction of these narrowly endemic species has been greatly accelerated in recent years by the introduction of alien species into island environments. The plight of these extraordinary sets of species has largely been ignored because attention has focused on the devastation of forests in South America and Asia. However, setting aside even small reserves would serve to protect many species on islands from extinction. The microcosmal nature of species swarms makes such efforts at least feasible. However, without immediate action, few of the world's most spectacular radiations will survive far into the next millennium.

See Also the Following Articles

COEVOLUTION • DARWIN, CHARLES • DISPERSAL BIOGEOGRAPHY • DIVERSITY, MOLECULAR LEVEL • ENDEMISM • HABITAT AND NICHE, CONCEPT OF • POPULATION GENETICS

Bibliography

Adler, G. H., and Dudley, R. (1994). Butterfly biogeography and endemism on tropical Pacific islands. *Biol. J. Linnean Soc.* **52**, 151–162.

Albertson, R. C., Markert, J. A., Danley, P. D., and Kocher, T. D. (1999). Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. *Proc. Natl. Acad. Sci. USA* **96**, 5107–5110.

Avise, J. C. (1991). Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annu. Rev. Genet.* **25**, 45–69.

Baldwin, B. (1997). Adaptive radiation of the Hawaiian silversword alliance: Congruence and conflict of phylogenetic evidence from molecular and non-molecular investigations. In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K. J. Sytsma, Eds.), pp. 103–128. Cambridge Univ. Press, Cambridge, UK.

Baldwin, B., and Robichaux, R. (1995). Historical biogeography and ecology of the Hawaiian silversword alliance (Compositae): New molecular phylogenetic perspectives. In *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago* (W. L. Wagner and V. A. Funk, Eds.), pp. 259–287. Smithsonian Institution Press, Washington, DC.

Baldwin, B. G., and Sanderson, M. J. (1998). Age and rate of diversification of the Hawaiian silversword alliance (Compositae). *Proc. Natl. Acad. Sci. USA* **95**, 9402–9406.

Barton, N. H. (1996). Natural selection and random genetic drift as causes of evolution on islands. *Philos. Trans. R. Soc. London* **351**, 785–795.

Berenbaum, M., and Feeny, P. P. (1981). Toxicity of furanocoumarins to swallowtail butterflies: Escalation in a coevolutionary arms race? *Science* **212**, 927–929.

Berenbaum, M. R. Favret, C., and Schuler, M. A. (1996). On defining "key innovations" in an adaptive radiation: Cytochrome P450s and Papilionidae. *Am. Nat.* **148**(Suppl.), S139–S155.

Bernebaum, M. (1983). Coumarins and caterpillars: A case for coevolution. *Evolution* **37**, 163–179.

Boake, C. R. B., DeAngelis, M. P., and Andreadis, D. K. (1997). Is sexual selection and species recognition a continuum? Mating behavior of the stalk-eyed fly *Drosophila heteroneura. Proc. Natl. Acad. Sci. USA* **94**, 12442–12445.

Cameron, R. A. D., Cook, L. M., and Hallows, J. D. (1996). Land snails on Porto Santo: Adaptive and non-adaptive radiation. *Philos. Trans. R. Soc. London B.* **351**, 309–327.

Carr, G. D. (1985). Monograph of the Hawaiian Madiinae (Asteraceae): *Argyroxyphium, Dubautia,* and *Wilkesia. Allertonia* **4**, 1–123.

Carr, G. D. (1999). A reassessment of *Dubautia* (Asteraceae: Heliantheae-Madiinae) on Kauai. *Pacific Sci.* **53**, 144–158.

Carson, H. L. (1986). Sexual selection and speciation. In *Evolutionary Processes and Theory* (S. Karlin and E. Nevo, Eds.), pp. 93–107. Academic Press, New York.

Carson, H. L. (1989). Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura. Evolution* **43**, 190–203.

Carson, H. L. (1990). Increased genetic variance after a population bottleneck. *Trends Ecol. Evol.* **5**, 228–230.

Carson, H. L. and Templeton, A. R. (1984). Genetic revolutions in relation to speciation phenomena: The founding of new populations. *Annu. Rev. Ecol. Syst.* **15**, 97–131.

Cox, G. W. and Ricklefs, R. E. (1977). Species diversity, ecological release, and community structuring in Caribbean land bird faunas. *Oikos* **29**, 60–66.

Darlington, P. J., Jr. (1957). *Zoogeography: The Geographical Distribution of Animals*. Wiley, New York.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. John Murray, London.

DeSalle, R. and Giddings, L. V. (1986). Discordance of nuclear and mitochondrial DNA phylogenies in Hawaiian *Drosophila. Proc. Natl. Acad. Sci. USA* **83**, 6902–6906.

DeSalle, R., Brower, A. V. Z., Baker, R., and Remsen, J. (1997). A hierarchical view of the Hawaiian Drosophilidae. *Pacific Sci.* **51**, 462–474.

Dommergues, J.-L., Laurin, B., and Meister, C. (1996). Evolution of ammonoid morphospace during the early Jurassic radiation. *Paleobiology* **22**, 219.

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants: A study in coevolution. *Evolution* **18**, 586–608.

Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281**, 555–559.

Farrell, B. D., and Mitter, C. (1994). Adaptive radiation in insects and plants: Time and opportunity. *Am. Zool.* **34**, 57–69.

Farrell, B. D., Dussourd, D. E., and Mitter, C. (1991). Escalation of plant defense: Do latex and resin canals spur plant diversification? *Am. Nat.* **138**, 881–900.

Freeland, J. R., and Boag, P. T. (1999). The mitochondrial and nuclear genetic homogeneity of the phenotypically diverse Darwin's ground finches. *Evolution*, **53**, 1553–1563.

Gillespie, R. G., Croom, H. B., and Hasty, G. L. (1997). Phylogenetic relationships and adaptive shifts among major clades of tetragnathid spiders in Hawaii. *Pacific Sci.* **51**, 380–394.

Gillespie, R. G., Rivera, M., and Garb, J. E. (1998). Sun, surf and spiders: Taxonomy and phylogeography of Hawaiian Araneae. In *Proceedings of the 17th European Colloquium of Arachnology, Edinburgh 1997* (P. A. Selden, Ed.), pp. 41–51. British Arachnological Society, Burnham Beeches, Bucks, United Kingdom.

Gittenberger, E. (1991). What about non-adaptive radiation? *Biol. J. Linnean Soc.* **43**, 263–272.

Givnish, T. (1997). Adaptive radiation and molecular systematics: Issues and approaches. In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K. J. Sytsma, Eds.), pp. 1–54. Cambridge Univ. Press, Cambridge, UK.

Givnish, T., and Sytsma, K. J. (1997). *Molecular Evolution and Adaptive Radiation*. Cambridge Univ. Press, Cambridge, UK.

Goodnight, C. J. (1988). Epistasis and the effect of founder events on the additive genetic variance. *Evolution* **42**, 441–454.

Grant, P. R. (1986). *Ecology and Evolution of Darwin's Finches*. Princeton Univ. Press, Princeton, NJ.

Hardy, D. E. (1965). Diptera: Cyclorrhapha II, Series Schizophora, Section Acalypterae. I. Family Drosophilidae. *Insects Hawaii* **12**, 1–814.

Hardy, D. E., and Kaneshiro, K. Y. (1981). Drosophilidae of Pacific Oceania. In *The Genetics and Biology of Drosophila* (M. Ashburner, H. L. Carson, and J. N. Thompson, Eds.), pp. 309–348. Academic Press, New York.

Harrison, R. G. (1993). *Hybrid Zones and the Evolutionary Process*. Oxford Univ. Press, Oxford.

Huxley, J. (1942). *Evolution: The Modern Synthesis*. Harper & Brother, New York.

James, H. F., and Olson, S. L. (1991). *Descriptions of thirty-two new species of birds from the Hawaiian Islands: Part II. Passeriformes. Ornith, Monograph No. 46*. American Ornithologists' Union, Washington, DC.

Johnson, M. S., Murray, J., and Clarke, B. (1993). The ecological genetics and adaptive radiation of Partula on Moorea. *Oxford Surv. Evol. Biol.* **9**, 167–238.

Kambysellis, M. P. (1993). Ultrastructural diversity in the egg chorion of Hawaiian *Drosophila* and *Scaptomyza:* Ecological and phylogenetic considerations. *Int. J. Insect Morphol. Embryol.* **22**, 417–446.

Kambysellis, M. P., and Craddock, E. M. (1997). Ecological and reproductive shifts in the diversification of the endemic Hawaiian *Drosophila.* In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K. J. Sytsma, Eds.), pp. 475–509. Cambridge Univ. Press, Cambridge, UK.

Kaneshiro, K. Y. (1983). Sexual selection and direction of evolution in the biosystematics of the Hawaiian Drosophilidae. *Annu. Rev. Entomol.* **28**, 161–178.

Kaneshiro, K. Y. (1988). Speciation in the Hawaiian *Drosophila. Bioscience* **38**, 258–263.

Kaneshiro, K. Y. (1989). The dynamics of sexual selection and founder effects in species formation. In *Genetics, Speciation and the Founder*

Principle (L. V. Giddings, K. Y. Kaneshiro, and W. W. Anderson, Eds.), pp. 279–296. Oxford Univ. Press, Oxford.

Klicka, J., and Zink, R. M. (1997). The importance of recent ice ages in speciation: A failed paradigm. *Science* **277**, 1666–1669.

Lack, D. (1947). *Darwin's Finches*. Cambridge Univ. Press, Cambridge, UK.

Lehmann, U. (1981). *The Ammonites*. Cambridge Univ. Press, Cambridge, UK.

Liebherr, J. K., and Polhemus, D. A. (1997). R. C. L. Perkins: 100 years of Hawaiian entomology. *Pacific Sci.* **51**, 343–355.

Liem, K. F. (1973). Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Syst. Zool.* **22**, 425–441.

Losos, J. B. (1992). The evolution of convergent structure in Caribbean *Anolis* communities. *Syst. Biol.* **41**, 403–420.

Losos, J. B., Jackman, T. R., Larson, A., deQueiroz, K., and Rodriquez-Scheltino, L. (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science* **279**, 2115–2118.

MacArthur, R. H., and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton Univ. Press, Princeton, NJ.

Margulis, L., and Fester, R. (1991). *Symbioses as a Source of Evolutionary Innovation*. MIT Press, Cambridge, MA.

McCune, A. R. (1997). How fast is speciation? Molecular, geological, and phylogenetic evidence from adaptive radiation in fishes. In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K. J. Sytsma, Eds.), pp. 585–610. Cambridge Univ. Press, Cambridge, UK.

McCune, A. R., and Lovejoy, N. R. (1998). The relative rates of sympatric and allopatric speciation in fishes. In *Endless Forms: Species and Speciation* (D. J. Howard and S. H. Berlocher, Eds.), pp. 171–185. Oxford Univ. Press, Oxford.

Meyer, A. (1993). Phylogenetic relationships and evolutionary processes in East African cichlids. *Trends Ecol. Evol.* **8**, 279–284.

Meyer, A., Montero, C. M., and Spreinat, A. (1996). Molecular phylogenetic inferences about the evolutionary history of the East African cichlid fish radiations. In *International Decade of East African Lakes: The Limnology, Climatology and Paleoclimatology of the East African Lakes* (T. Johnson and E. Odada, Eds.), pp. 303–323. Gordon & Breach, New York.

Mitter, C., Farrell, B., and Wiegmann, B. (1988). The phylogenetic study of adaptive zones: Has phytophagy promoted insect diversification. *Am. Nat.* **132**, 107–128.

Moran, P., and Kornfield, I. (1993). Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei: Cichlidae) of Lake Malawi. *Mol. Biol. Evol.* **10**, 1015–1029.

Murray, J., Clarke, B., and Johnson, M. S. (1993). Adaptive radiation and community structure of *Partula* on Moorea. *Proc. R. Soc. London B* **254**, 205–211.

Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution* **29**, 1–10.

Nevo, E., Filippucci, M. G., and Beiles, A. (1994). Chromosomal speciation and adaptive radiation of mole rats in Asia Minor correlated with increased ecological stress. *Proc. Natl. Acad. Sci. USA* **91**, 8160–8164.

Norris, R. D. (1996). Symbiosis as an evolutionary innovation in the radiation of Paleocene planktic foraminifera. *Paleobiology* **22**, 461–480.

Osborn, H. F. (1902). The law of adaptive radiation. *Am. Nat.* **36**, 353–363.

Otte, D. (1994). *The Crickets of Hawaii: Origin, Systematics, and Evolution*. The Orthopterists' Society, Academy of Natural Sciences, Philadelphia.

Otto, S. P., and Whitlock, M. C. (1997). The probability of fixation in populations of changing size. *Genetics* **146**, 723–733.

Petren, K., Grant, B. R., and Grant, P. R. (1999). A phylogeny of Darwin's finches based on microsatellite DNA length variation. *Proc. R. Soc. London B* **266**, 321–329.

Rainey, P. B., and Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69–72.

Raven, P. H., Evert, R. F., and Eichhorn, S. E. (1992). *Biology of Plants*, 5th ed. Worth, New York.

Roderick, G. K., and Gillespie, R. G. (1998). Speciation and phylogeography of Hawaiian terrestrial arthropods. *Mol. Ecol.* **7**, 519–531.

Sato, A., O'hUigin, C., Figueroa, F., Grant, P. R., Grant, B. R., Tichy, H., and Klein, J. (1999). Phylogeny of Darwin's finches as revealed by mtDNA sequences. *Proc. Natl. Acad. Sci. USA* **96**, 5101–5106.

Schluter, D. (1994). Experimental evidence that competition promotes divergence in adaptive radiation. *Science* **266**, 798–801.

Schluter, D. (1996). Ecological causes of adaptive radiation. *Am. Nat.* **148**, S40–S64.

Schluter, D. (1998). Ecological causes of speciation. In *Endless Forms: Species and Speciation* (D. J. Howard and S. H. Berlocher, Eds.), pp. 114–129. Oxford Univ. Press, Oxford.

Schluter, D., and McPhail, J. D. (1993). Character displacement and replicate adaptive radiation. *Trends Ecol. Evol.* **8**, 197–200.

Schneider, C., and Moritz, C. (1999). Rainforest refugia and evolution in Australia's wet tropics. *Proc. R. Soc. London* **266**, 191–196.

Shaw, K. L. (1996). Sequential radiations and patterns of speciation in the Hawaiian cricket genus *Laupala* inferred from DNA sequences. *Evolution* **50**, 237–255.

Shields, O. (1993). The origin and initial radiation of angiosperms in relation to Anthophytes. *Palaeobotanist* **42**, 157–185.

Simpson, G. G. (1953). *The Major Features of Evolution*. Columbia Univ. Press, New York.

Slatkin, M. (1996). In defense of founder-flush theories of speciation. *Am. Nat.* **147**, 493–505.

Sloan, R. E., Rigby, J. K., Jr., Van Valen, L. M., and Gabriel, D. (1986). Gradual dinosaur extinction and simultaneous ungulate radiation in the Hell Creek Formation. *Science* **232**, 629–633.

Smith, A. (1984). *Palaeobiology*. Allen & Unwin, London.

Springer, M. S., Kirsch, J. A. W., and Case, J. A. (1997). The chronicle of marsupial evolution. In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K. J. Sytsma, Eds.), pp. 129–161. Cambridge Univ. Press, Cambridge, UK.

Stauffer, J. R., Jr., Bowers, N. J., Kellogg, K. A., and McKaye, K. R. (1997). A revision of the blue-black *Pseudotropheus* zebra (Teleostei: Cichlidae) complex from Lake Malawi, Africa, with a description of a new genus and ten new species. *Proc. Acad. Nat. Sci. Philadelphia* **148**, 189–230.

Sturmbauer, C. (1998). Explosive speciation in cichlid fishes of the African Great Lakes: A dynamic model of adaptive radiation. *J. Fish Biol.* **53**(Suppl. A), 18–36.

Tarr, C. L., and Fleischer, R. C. (1995). Evolutionary relationships of the Hawaiian honeycreepers (Aves: Drepanidae). In *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago* (W. L. Wagner and V. A. Funk, Eds.), pp. 147–159. Smithsonian Institution Press, Washington, DC.

Taylor, E. B., McPhail, J. D., and Schluter, D. (1997). History of ecological selection in sticklebacks: uniting experimental and phylogenetic approaches. In *Molecular Evolution and Adaptive Radiation* (T. J. Givnish and K J. Sytsma, Eds.). pp. 511–534. Cambridge University Press, Cambridge.

Templeton, A. R. (1980). The theory of speciation via the founder principle. *Genetics* **91**, 1011–1038.

Vermeij, G. J. (1995). Economics, volcanoes, and Phaerozoic revolutions. *Paleobiology* **21**, 125–152.

Wagner, W. L., and Funk, V. A. (1995). *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago*. Smithsonian Institution Press, Washington, DC.

Whittaker, R. J. (1998). *Island Biogeography: Ecology, Evolution, and Conservation*. Oxford Univ. Press, Oxford.

Wilson, E. O. (1961). The nature of the taxon cycle in the Melanesian ant fauna. *Am. Nat.* **95**, 169–193.

AESTHETIC FACTORS

Gordon H. Orians
University of Washington

GLOSSARY

aesthetics The field of investigation, from philosophical and psychological perspectives, that attempts to discover the rules and principles that govern the sense of beauty and ugliness.

beauty The qualities of a perceived or imagined object whereby it evokes feelings of admiring pleasure.

cognition Act or process of knowing and understanding.

habitat Type of environment in which an organism lives, grows, and reproduces.

phobia Obsessive fear or dread of an object or situation.

symmetry The correspondence in size, form, and arrangement of parts on opposite sides of a two- or three-dimensional object.

AESTHETIC FACTORS are those characteristics of a given object or situation that evoke a certain emotional response, either a sense of beauty, attractiveness, pleasure, symmetry, order, and so on, or, conversely, of ugliness, disorder, menace, disgust, or the like. Generally speaking, the aesthetic preferences that humans display in response to their environment, in such contexts as mate choice, food patterns, and habitat selection, have been shaped by evolutionary experience and reflect suitable solutions for survival and reproductive success.

I. INTRODUCTION

Humans have strong emotional responses to living organisms and to natural and human-modified environments. Depending on the circumstances and the needs of an individual, nature may evoke feelings of awe, respect, fear, loathing, longing, nostalgia, excitement, challenge, and belonging. These powerful emotions influence how we respond to nature, how we attempt to manipulate it, and why we care about it.

Strong emotional responses in all organisms, including humans, evolved because, on average, they increased fitness, that is, they improved the survival and reproductive success of the individuals that expressed them. For example, those of our ancestors who did not enjoy food and sex certainly were more poorly represented in subsequent generations than those who did enjoy and hence sought out food and sexual partners. Similarly, individuals who chose ecologically inferior environments in which to live should have been less well represented genetically in subsequent generations than individuals who made better habitat choices.

However, this truism does little more than establish general guidelines for investigating why human emotional responses to components of the environment are so varied. Nor does it take us very far in improving our understanding of how different emotional responses to varied environmental components resulted in fitness-improving behaviors. More specific theories and analyses are needed.

II. EVOLUTIONARY APPROACHES TO ENVIRONMENTAL AESTHETICS

Why human emotional responses to nature would have influenced fitness is obvious. Our ancestors lived in environments devoid of modern conveniences. Their health, survival, and reproductive success depended on their ability to obtain and use environmental information wisely. They had to know how to interpret signals from the physical and biological components of the envirnoment, and to adjust their behavior in response to them. They needed to evaluate habitats and the resources provided over space and time, and to adjust their use of habitats accordingly. Our ancestors also needed to be able to detect the presence of other humans, their effects on resources, and the direct dangers they posed or the help they might provide.

Making these evaluations and decisions was and still is difficult because the environment provides a far richer array of information than can possibly be assimilated and synthesized. Adaptive behavior requires selective attention to components of the environment that strongly influence fitness. In fact, the complex behavior of organisms would be impossible in the absence of neural filters that emphasize or deemphasize aspects of information emanating from the environment. These preselection filters embody evolutionarily stored knowledge about the world that enables us to construct hypotheses capable of describing and understanding the external world. Extensive experiments on artificial intelligence have clearly shown that the rapid and efficient learning of a language performed by nearly every human child is quite impossible in the absence of preformed neural structures (Dennett, 1995).

These prefilters evidently express themselves in all organisms as a sense of aesthetics and some form of logical analysis (Marchetti, 1998). In other words, aesthetic responses, rather than being a recently acquired capacity of little evolutionary significance, are fundamental to the ways in which organisms learn about the world and adapt to it. The English word *aesthetic* is derived from the adjectival form of the Greek *aisthanomai,* which means "to perceive." Thus, "aesthetic pleasure" means literally "pleasure associated with or deriving from perception." The central problem in aesthetics is to explain why pleasure and disgust have evolved to be so strongly associated with perception and recognition of certain kinds of objects.

The evolution of preselecting filters is reflected in what is called *biologically prepared learning* (Seligman, 1970). Biologically prepared learning theory asserts that evolution has predisposed humans, as well as individuals of other species, to learn easily and quickly, and to retain associations or responses that foster survival when certain objects or situations are encountered. An important corollary is that even though modern societies have greatly reduced the real danger posed by the objects of fears and phobias, fear and avoidance responses may nonetheless persist because selection against those responses is weak (Ulrich, 1993; Wilson, 1984). Biologically prepared learning in humans should be evident only for stimuli that have had significant influences on survival and reproductive success during our evolutionary history. Preselected responses should arise quickly when the appropriate conditioning stimuli appear, and these responses should be unusually resistant to extinction or forgetting. Adaptive responses to natural stimuli do not necessarily appear spontaneously or in the absence of learning, but they may.

Although aesthetic inquiries have been carried out since the middle of the eighteenth century, attention has been directed almost exclusively to responses to human creations rather than to nature. The neglect of nature is a by-product of the view that has dominated Western thinking for centuries—namely, that aesthetic experiences are molded primarily by cultural symbols and art forms. As a result, the study of aesthetics has been viewed as the domain of artists and philosophers. Attempts to explore the biophysical bases of aesthetic responses to environments were regarded, and still are by some people, as both futile and ideologically dangerous.

Culture and learning clearly exert strong influences on the ways humans perceive and respond to environmental information and they have important impacts on the symbolisms we attach to natural objects (Appleton, 1990; Schama, 1995). But attempts to understand and interpret human aesthetic responses to environmental features without asking why they evolved have failed.

For example, people who studied the concept of beauty from a nonbiological perspective assumed that beauty was an intrinsic property of objects. They therefore looked for, and expected to find, correlations between the characteristics of objects and human aesthetic responses to them. This attempt largely failed because, as an evolutionary perspective immediately suggests, beauty is not an intrinsic property of the objects that we call beautiful. Rather, it is the product of interactions between traits of objects and the human nervous sytem that evolved so that we regard as beautiful those objects having properties that, if positively attended to, result in improved performance in some aspect of living (Appleton, 1975). Conversely, we regard as ugly those objects that should be avoided or destroyed. In essence, an evolutionary perspective suggests that concepts such as beauty and ugliness should be viewed from a functional rather than a structural perspective. In other words, emotional responses are best studied by asking "How did these responses help us solve problems?"

III. ENVIRONMENTAL INFORMATION, PROBLEM SOLVING, AND SURVIVAL

Individuals of all species use information to make decisions that enhance their survival chances. The array of information presented to an individual has both components that are highly relevant to survival and reproductive success, and components that can be ignored with few or no adverse consequences. In addition, which components are important varies with time, location, and the needs of the individual.

Progress in the study of environmental aesthetics and problem solving has been aided by the development of systems for classifying environmental information into categories that roughly correspond to the kinds of decisions that organisms must make (Heerwagen and Orians, 1993). The basic problems that people (and other animals) must solve are: (1) protecting themselves from being injured or killed by other people or dangerous animals; (2) avoiding being attacked by parasites and disease-causing organisms that may be acquired from people or from other sources; (3) protecting themselves from adverse physical conditions; (4) acquiring enough of the right kinds of food; and (5) choosing good-quality associates for reproduction, foraging, protection, and achieving higher social status.

One useful classification divides information into categories based on the type of objects being identified. Some of these objects, such as food, water, and refuges, are resources. Gathering information about these resources should be pleasureable and exciting. Other objects, such as potentially dangerous animals, competitors, human enemies, and hazardous physical objects, pose dangers. Some of them are life-threatening, but others, although potentially dangerous, can be studied and overcome. Gathering information about them should be accompanied by feelings of fear, anxiety, and apprehension rather than pleasure (Table I).

Another classification of environmental information is based on the time frame over which the information is relevant. Although time is a continuous variable, humans long ago discovered that it is heuristically useful to divide time into categories. From the perspective of the study of environmental aesthetics, time is most usefully divided into categories that correspond to the time frames over which decisions about them matter (Table II). Some environmental information signals events of temporary significance. Still other information signals seasonal changes that are associated with shifts in the types and locations of resources that will be available in the near future. Still other information signals features of the environment, such as status of vegetation and courses of rivers, that change slowly over periods of years or decades. Finally, other information arrives from objects that, measured in terms of human lifetimes, are permanent.

A third classification is based on the sensory mode in which the information arrives. What information can be perceived obviously depends on the sensory capacity

TABLE I

Classification of Information by Type of Object

Category	Examples	Decisions affected
Inanimate objects		
Stationary	Landforms, glaciers	Bases of operations
Mobile	Rivers, ocean surface	Timing of foraging and exploratory activities
Animate objects		
Resources	Food items, fiber, medicine sources	Types, locations, and timing of foraging trips
Dangerous	Predators, large prey, venomous plants and animals, disease-causing or disease-vectoring animals	Types, locations, and timing of foraging trips

TABLE II

Classification of Information in Terms of the Time Frame of Its Relevance[a]

Category	Examples	Decisions affected
Short term (minutes to hours)	Weather changes (thunderclouds, wind)	Seeking shelter, initiating outdoor activities
	Appearance of dangerous animals, enemies	Immediate defensive actions
	Appearance of valuable prey	Immediate hunting activities
	Illumination changes	Moves to appropriate locations for spending the night
Seasonal	Day-length changes, vegetative growth, flowering, precipitation changes	Shifts of hunting sites, planting and harvesting of crops
Multiyear changes	Vegetation succession, erosional changes (river meanders, lake sedimentation)	Shifts of hunting sites, movement of villages
Long term (decades to centuries)	Topography	Development of traditions

[a] After Orians (1998).

of an organism and its neural processing system. The first, and still the most important sensory capacity in the living world as a whole, is chemical sensitivity. Molecules arrive by diffusion, augmented by movements of the medium. All chemicals are sensed when they come into contact with an organism's body, but, based on how fast they are carried by air currents, we distinguish verbally between the senses of taste and smell. The chemical knowledge possessed by a single bacterial cell exceeds what the most sophisticated organic chemist can measure. And because the sense of taste is based on chemical reactions, noses can be scaled down almost to the size of molecules.

Light sensitivity also evolved early in life's history on Earth. Bacteria evolved both photosensitivity—light-sensitive spots—and photosynthesis. The former vastly increased the distance over which objects could be detected; the latter generated nearly all of the energy in today's biosphere. Because light travels in straight lines, at the speed of light, vision potentially provides accurate and time-specific information about shapes of objects and their location in space. Sound has similar properties to light, but location of objects is not given precisely. Animals have evolved complex ways of inferring locations of sources of sounds by comparing events at the two ears.

Nervous systems evolved in part because, by building up parallel computing units, they enabled organisms to speed up and amplify responses to sensory inputs. For motile organisms, speed of response is often extremely important. Parallel computing units also are the units in which preselected filters are stored.

Responses to environmental information are complex because the significance of objects and events varies with the location and needs of observers. An approaching storm may be welcomed by a farmer whose crops are wilting from lack of rain, but despised by members of a family enjoying their first picnic of the summer. A vivid sunset enjoyed by people relaxing on the deck of their home may bring fear to a person walking on an African savanna too far from home to reach it before dangerous nocturnal predators become active. For these reasons, simple correlations between features of objects and aesthetic evaluations are unlikely ever to be discovered, although correlations are likely between types of objects, contents, and aesthetic responses.

Three functional concepts—prospect, refuge, and hazard—have guided recent approaches to the role of environmental aesthetics in problem solving (Appleton, 1975). *Prospect* refers to the ability of an individual to gather information about an environment with which to evaluate its characteristics and decide how to use it. Environments high in prospect offer rich opportunities for evaluation; environments low in prospect offer fewer opportunities. *Refuge* refers to the degree to which an environment provides security for an individual from negative agents while exploration and information gathering are taking place. *Hazard* refers to the dangers to which an individual would be exposed during information-gathering activities.

These concepts have been applied primarily to the initial evaluation and exploration of unfamiliar environments, but when combined with more recent developments in evaluating environmental information, they are readily applied and extended to a rich array of circumstances in which people must solve problems.

Emotional responses that assist in solving these problems necessarily include both positive and negative components. Because they are simpler and easier to study, negative responses are discussed first.

IV. NEGATIVE EMOTIONAL RESPONSES

Many people have strong fearful reactions—phobias—to a wide variety of objects or situations (snakes, spiders, heights, closed spaces, open spaces, blood). The set of objects that evoke fearful responses appears to be the same in all industrialized societies for which data are available (Ulrich, 1993), but few data exist on the prevalence of phobias in developing nations and nontechnological societies.

The objects and situations that evoke fearful responses are generally ones that have been associated with threatening situations during human evolutionary history. Precipitous cliffs are dangerous if approached too closely. Closed spaces offer few escape routes; people in wide open spaces are vulnerable to attacks by enemies and dangerous predators. Venomous and predatory animals have been significant sources of human injury and mortality for many millennia. Interestingly, even though they account for trivial numbers of injuries and deaths in modern societies, spiders and snakes are the objects of the most prevalent phobias in Western societies today.

Research on human twins has provided convincing evidence that genetic factors play major roles in a wide range of human traits, including animal phobias and fear of open spaces. Imaginative research, especially by psychologists in Sweden and Norway, shows that biologically prepared learning plays a significant role in the acquisition and retention of phobias (see Ulrich, 1993, for an overview of experimental results). The most informative experiments employ a Pavlovian conditioning approach to compare the development of defensive and aversive conditioned responses (learned through repeated exposure or experience) to slides of fear-relevant and fear-irrelevant or neutral stimuli. Defensive responses are assessed by recording autonomic nervous system indicators such as skin conductance and heart rate.

Experimenters condition initial defensive responses by showing either fear-relevant (snakes or spiders) or neutral stimuli (geometric figures) paired with an aversive stimulus (the unconditioned stimulus), usually an electric shock intended to mimic a bite. This phase of the experiments makes it possible to compare the speed and strength of acquisition of defensive/aversive responses to fear-relevant and neutral stimuli.

Following an initial acquisition phase, the same stimuli are presented many times without reinforcement of an electric shock. This "extinction" phase allows comparison of the rate of extinction of the defensive/aversive response acquired earlier. The general result is that conditioned responses are usually, but not always, acquired more quickly, but that responses to snakes and spiders are reliably more resistant to extinction than responses to neutral stimuli. These responses cannot be the result of prior cultural reinforcement, because conditioned aversive responses to familiar modern dangerous stimuli, such as hand guns and frayed electrical wires, extinguish more rapidly than conditioned responses to snakes and spiders. In addition, aversive responses to fear-relevant natural stimuli can be elicited merely by telling a person that a shock will be administered. Aversive responses to fear-irrelevant natural stimuli cannot be elicited in this manner.

People also acquire much more persistent defensive reactions when watching an experimenter's reactions to fear-relevant than to fear-irrelevant stimuli. Similar results have been obtained by exposing rhesus monkeys to fear-relevant (toy snakes and crocodiles) and fear-irrelevant stimuli (toy rabbits).

Even more striking are the results of "backwardmasking" experiments, in which slides are displayed subliminally (for 15–30 milliseconds) before being "masked" by a slide of another stimulus or setting. Even though the subjects are not consciously aware of having seen the stimulus slide, presentations of natural settings that contain snakes or spiders elicit strong aversive/defensive reactions in nonphobic persons. If a previous conditioning has already occurred, a masked subliminal presentation is sufficient to elicit defensive responses to the feared stimulus (Öhman and Soares, 1994).

Thus, a rich array of experimental results demonstrates that aversive responses develop more rapidly and persist longer to fear-relevant than to fear-irrelevant natural stimuli. In addition, defensive/aversive responses can develop to natural threat stimuli even though subjects are unaware that they have seen them. Such responses do not develop to neutral or fear-irrelevant stimuli. The patterns of these responses demonstrate the powerful role of biologically prepared learning, and they make adaptive sense.

V. POSITIVE EMOTIONAL RESPONSES

Research and theory are less well developed for positive emotional responses to environmental stimuli than for negative ones. No studies of the genetics of positive emotional responses have been carried out, and positive conditioning studies are relatively rare. This is in part due to the fact that positive Pavlovian conditioning experiments are typically more difficult to conduct than aversive conditioning experiments. In addition, positive emotional responses are very complex and difficult to quantify.

A. Symmetry and Beauty

Extensive information has been gathered on emotional responses to the symmetry of objects. A rich body of literature demonstrates that humans find symmetrical objects, including abstract patterns, woody plants, and human bodies—especially faces—aesthetically pleasing. Moreover, sexual preferences for symmetrical facial features appear to be similar across human cultures. Other vertebrates also are sensitive to very small asymmetries (see Møller and Swaddle, 1997, for a review of the human and animal literature). Conversely, asymmetries typically evoke negative emotional responses, the most extreme of which is selective infanticide of asymmetrical newborn babies in many human cultures.

There are several nonmutually exclusive reasons why positive responses to symmetry are adaptive. One is that symmetrical animals perform better physically than asymmetrical animals, just as symmetrical objects, such as bows, arrows, axes, boats, autos, and airplanes, function better than asymmetrical ones. Asymmetrical individuals may have been exposed to severely stressful environments that disrupted normal development, thereby rendering them less functionally adequate. For example, alcoholic mothers give birth to children with greater developmental asymmetries than do mothers who consumed less alcohol. Thus, it is not surprising that choices of objects to use, the design of objects, and choices of social partners have evolved to favor use of and association with symmetrical objects.

Second, asymmetrical individuals are likely to have genetic defects that could be passed on to offspring. For example, humans suffering from trisomy-21 (Down syndrome) have noticeable asymmetries due to skeletal abnormalities. Greater asymmetry in dental traits is associated with genetic disorders. Therefore, avoiding asymmetrical individuals as mates would be likely to impove the quality of an individual's offspring.

Third, because diseases can cause asymmetrical development, asymmetrical individuals may be avoided because they may currently harbor communicable disease-causing microorganisms. Good symmetry signals a long history of good health!

B. Habitat Selection

Habitat selection is a vital decision in the lives of all organisms. When selecting a habitat an organism responds as *if* it understood the significance of objects, sounds, and odors for its future survival and reproductive success. Initial responses typically are emotional feelings that lead to rejection, exploration, or a certain use of the environment. Because the strength of these responses is a key to immediate decisions about where to settle and what to do there, the nature of the emotional responses evoked by habitats should evolve to be positively correlated with the expected survival and reproductive success of an organism in those habitats. That is, "good" habitats should evoke strong positive responses; "poor" habitats should evoke weak or even negative responses.

Issues of spatial scale are important during habitat selection. At small spatial scales, the primary decisions involve food and shelter. The key decisions are whether to accept or reject a specific food item or shelter. At medium spatial scales, the key decisions are in which activity to engage and in which patches in the environment those activities will be carried out. At still larger spatial scales, decisions center on whether to initiate major changes in where to carry out activities, that is, to migrate or to shift the base of operations.

Theoretical explorations of these decisions typically include the development of "ideal" models, that is, models that assume perfect knowledge on the part of the decision maker. These models specify the best that an organism can do in solving a problem (e.g., obtaining the most food per unit hunting time). They are unrealistic because knowledge is always incomplete, but they are valuable for assessing the marginal value of additional information and for determining how well an animal using rather simple "rules of thumb" would perform compared to an "ideal" animal. Many of these rules of thumb include aesthetic reponses that guide decisions. It is useful to begin with a consideration of medium-scale responses and then turn to microscale responses.

Habitat selection, viewed at medium spatial scales, has served as a perspective for a number of studies on human aesthetic responses to landscape features (Heerwagen and Orians, 1993). Habitats occupied by humans during most of our evolutionary history rarely provided resources that were reliable long enough to enable permanent occupation of sites. Frequent moves through the landscape were the rule even though traditional sites might be revisited on an annual basis. Because relatively few generations have passed since humans started to live in mechanized and urban environments, evolutionarily based response patterns of humans to landscapes are unlikely to have been substantially modified since the rise of industrialized, urban societies.

Human responses to environmental cues vary with a person's age, social status, and physiological state. Nevertheless, positive responses to indicators of the presence of food, water, shelter, and protection from predators are general. So are negative responses to potential hazards, such as inclement weather, fire, dangerous predators, and barriers to movement. Although no direct evidence yet exists for genetic influences on these responses, a number of evolutionary hypotheses have generated predictions, some of which have been tested experimentally.

One approach is based on the fact that *Homo sapiens* evolved in African savannas and only recently has invaded other continents. Therefore, landscape features and tree shapes characteristic of high-quality African savannas are expected to be especially attractive to humans today. This hypothesis has been tested by determining the responses of people to tree shapes and by examining the features of "aesthetic environments," that is, those environments, such as parks and gardens, that are designed to make them attractive (Orians, 1986).

The shapes of trees that dominate savannas are good predictors of the resource-providing capacities of those environments. Therefore, people evolved to find more pleasing the shapes of trees that were prominent in environments that provided the highest-quality resources rather than shapes of trees that dominated poor-quality habitats. Trees that grow in the highest-quality African savannas have canopies that are broader than they are tall, trunks that bifurcate close to the ground, and layered canopies. People on three continents preferred Kenyan *Acacia tortilis* trees that had highly or moderately layered canopies, lower trunks, and higher canopy width/tree height ratios than trees with narrow canopies and trunks that bifurcated higher above the ground (Heerwagen and Orians, 1993). Similarly, college students in Australia, Brazil, Canada, Israel, Japan, and the United States preferred trees with broad spreading crowns over conical and columnar trees (Sommer and Summit, 1996).

The changes that landscape architects recommend to their prospective customers are another source of data on human responses to landscapes. Humphrey Repton, an eighteenth-century British landscape architect, presented his clients with "before" and "after" drawings of their estates (Repton, 1907). Because he presumably wished to encourage rather than discourage potential clients, Repton changed landscapes by creating more savannlike scenes, by increasing visual access and penetrability of closed woods, by opening up distant views to the horizon (i.e., increased prospect), by adding refuges and cues signaling ease of movement, and by adding evidence of resource availability, particularly large mammals (Heerwagen and Orians, 1993).

Although a love of flowers is a pervasive human trait, it is not obvious why an omnivorous primate should take flowers to hospitals, bring them to dinner parties and house warmings, and annually spend billions of dollars on them. Nor is it obvious, given that flowers did not evolve their forms and colors to please us, why we should find them so aesthetically attractive. An evolutionary perspective suggests that flowers evoke strong positive feelings because they have long been associated with food resources. Because flowers precede fruits, flowering plants provide excellent cues to timing and locations of future resources. In addition, flowers may attract animals, especially birds, that are potential human prey. In species-rich environments, paying attention to flowering plants may particularly enhance resource acquisition abilities in the future. Until the nineteenth century, honey was the only natural source of sugar; beekeeping is an ancient human enterprise.

No studies have investigated which traits of flowers evoke strong positive feelings, but the obvious changes produced in many species of flowers by artificial selection—increased size and duplication of floral parts—result in flowers similar to those that historically produced large nectar rewards. Anectodal evidence suggests that strongly asymmetical flowers, which usually produce large nectar rewards, are generally more attractive than symmetrical ones, suggesting a possible exception to the general human preference for symmetrical objects. Future studies of human aesthetic repsonses to flowers are likely to provide interesting results.

C. Foods and Food Selection

Eating is essential for survival. Foraging animals make two major types of decisions: what items are acceptable as food, and which acceptable items are actually eaten during a foraging bout. Aesthetic responses are especially prominent in the first of these decisions. The strong emotional responses associated with foods are not surprising because eating requires voluntary ingestion of foreign objects. To eat, the body's generally aversive responses designed to protect health must be overcome; vomiting and diarrhea are the only postingestion defenses. These decisions are especially difficult for omnivores, who must ingest a variety of foods to achieve a balanced diet in varied environments where the array of available foods differs dramatically over space and time. Omnivores are expected to evolve only general food preselectors and should develop systems for classifying organisms with respect to their potential value as food.

People accept or reject foods for complex reasons (Table III). Direct sensory responses—tastes good or tastes bad—are important, but they are only one component of acceptability. People also make decisions on the basis of anticipated consequences, both physiological and social, of eating a type of food. In addition, particular foods may acquire ideational features that limit their acceptability to particular situations or cause them to be categorically rejected. Because of the variety of reasons that influence which foods are in the set of acceptables, human cuisines differ more than would be expected simply from knowing what edible resources are available in the environments in which social groups live (Rozin, 1996).

Despite the powerful idiosyncracies of human diets, a few generalizations have emerged. First, most items that evoke disgust are of animal origin. Plant parts rarely evoke disgust in any culture. On the other hand, "inappropriate" items are primarily vegetable in origin (Rozin and Fallon, 1981). Disgust may be an adaptation that deterred our ancestors from eating animal tissues, such as feces, rotting meat, and soft internal parts, all of which commonly harbor large numbers of potentially harmful microorganisms. Microorganisms have the ability to multiply rapidly, so there is no safe dose for ingesting them. Disgust may represent evolutionarily programmed intuitive microbiology.

Perhaps the most interesting feature of human cuisines is the stability of the major spices and sauces that characterize them. Humans are remarkably conservative in their food habits and are typically reluctant to try new foods or to abandon familiar ones. Traditional flavorings are high-priority culinary items; immigrant groups go to great lengths and expense to procure them in foreign settings. The deliberate manipulation of food by adding ingredients that reliably alter its taste is a uniquely human behavior. No other animals are known to do so. Which flavorings are used probably evolved in relation to what was available in the environment, but once established they are remarkably persistent. Certain flavorings probably signal that the food is safe. For example, the French refused to eat potatoes until Parmientier, an eighteenth-century French agriculturist, showed how to prepare them in familiar ways with familiar seasonings: butter, cheese, and herbs.

Cross-cultural similarities in which spices are used have evolved because spices inhibit or kill food-spoiling microorganisms. The most widely used spices all have strong antimicrobial properties (Billing and Sherman, 1998); mixes of them, which are common in many traditional recipes, are even more powerful. Not surprisingly, given that disease-causing organisms are more abundant in tropical than in temperate regions, the proportion of traditional recipes containing antimicro-

TABLE III

Reasons People Accept or Reject Foods[a]

Primary reason	Acceptance	Rejection
A. Sensory	Tastes good	Tastes bad
B. Anticipated Consequences		
Short-term physiological effects	No adverse reaction	Allergic reactions, gastrointestinal upsets
Long-term physiological effects	Societal concepts of "healthy foods"	Societal concepts of "unhealthy foods"
Social effects	Results in acceptance by group (smoking, social drinking)	Causes rejection by group (foods not eaten by that group, e.g., pigs in some Mideast cultures)
C. Ideational		
Appropriateness	Appropriate ritual foods	Inappropriate nonedible contaminants (soil, weeds)

[a] Modified from Rozin and Fallon (1981).

bial spices is inversely correlated with latitude. Alternative hypotheses—that spices provide micronutrients, disguise the taste and smell of spoiled foods, or increase perspiration and, thus, evaporative cooling—are not supported by the extensive data base (Billing and Sherman, 1998).

D. Restorative Responses

If aesthetic responses evolved because they enabled people to better solve life's problems, exposure to high-quality environments should be restorative, that is, it should reduce feelings of tension and stress. Stress reduction consistently emerges as one of the key benefits reported by recreationists in wilderness areas. Restoration from stress is also reported as a key benefit from time spent in urban parks with savannalike vegetation and water (Schroeder, 1989). Patients recovering from surgery in hospitals with either views of natural vegetation or simulated views that depict natural scenes with water recover more rapidly and have less postoperative anxiety than patients with no access to natural views or who are presented with simulations of abstract designs. Many studies have shown that even a brief exposure to nature, real or via photographs, leads to positive emotional feelings, reductions in stress, and better performance on demanding tasks (see Ulrich, 1995, for a review). Clearly, the affiliative responses people have to nature have important implications for the design of work and living spaces and healthcare facilities that are just beginning to be implemented.

VI. AESTHETICS AND BIODIVERSITY

Although people are strongly attracted to living organisms, it is less clear that the attractiveness of an environment is consistently positively correlated with the number of species in it. On the one hand, for example, the most highly evolved garden traditions—European formal gardens and Japanese gardens—are based on just a few species of woody plants. Landscape designers generally do not like the gardens of botanists because they are cluttered up with plants of too many species! Scenes of environments that contain a jumble of plants of many species receive low scores in psychological tests. Subjects report that they are too difficult to interpret; it is difficult to determine how to enter and use them.

On the other hand, people take great pleasure in finding as many species of birds as they can on a given day and from assembling "life lists" of species they have seen. Generally, seeing more species is better than seeing fewer species. Journeys of hundreds or even thousands of miles to see a rare species not on one's life list are not uncommon in today's mobile society. People are also powerfully attracted to the unusual—rare species, individuals outside the normal range of the species, or individuals present at unusual times of the year or in unusual habitats.

Because these familiar patterns of human behavior have been subjected to remarkably little formal study, we can only speculate about why they have evolved. That environments with intermediate levels of biological complexity should be preferred over both simpler and more complex environments makes sense, because the range of resources present in an environment and the ability to find and use those resources probably peak at intermediate levels of complexity. Simple environments have too few resources; complex ones have so many that choosing among them becomes difficult. Developing a suitable classification system to guide responses to the components of complex environments may be especially difficult. People may have evolved to respond to rare and unusual events because they provide new information about the state of the environment. Not all novel events are associated with something important, but it may be best to pay attention to them to find out if they are rather than to ignore the signals. Novel events may indicate that current patterns of use of the environment should be altered.

VII. AESTHETICS AND COGNITION

The human mind evolved into its current form long before the invention of agriculture and the dawn of the Industrial Revolution. It evolved its special characteristics in the service of our hunter-gatherer ancestors, who almost daily faced serious challenges from their physical, biological, and social environments. During the past century, many scientists found it difficult to explain why the demands of functioning in a preagricultural environment should have favored the evolution of the vast complexities of the human mind, which is capable of the many feats of which we are so proud. The apparent paradox arises because of a failure to appreciate that succeeding as a forager is probably more complicated than playing chess or doing calculus. To outwit nature, people need to use some form of intuitive scientific thinking, develop abstract conceptions, think about the future, compete in a rich social environment, and plan tactics. Aesthetic responses are basic components in all of these responses.

Mathematicians describe their search for important theorems as an aesthetic experience; they think of proven theorems as being beautiful. In fact, scientists in all fields regularly describe their models, experiments, and results in aesthetic terms. Thus, aesthetic sensibilities are imbedded deep in the human mind, having evolved in the service of choice of mates, habitats, and foods. An aesthetic sense functions as a holistic filter that helps the mind search efficiently for good solutions. It can do this because the filter has been molded by countless evolutionary experiences. Perhaps one of the most satisfying results of having a mind that was molded by natural selection is that we generally enjoy doing things that are good for us to do.

See Also the Following Articles

SOCIAL AND CULTURAL FACTORS • HISTORICAL AWARENESS OF BIODIVERSITY • HUMAN IMPACT ON BIODIVERSITY, OVERVIEW • LITERARY PERSPECTIVES ON BIODIVERSITY • TOURISM, ROLE OF

Bibliography

Appleton, J. (1975). *The Experience of Landscape.* John Wiley & Sons, New York.

Appleton, J. (1990). *The Symbolism of Habitat: An Interpretation of Landscape in the Arts.* University of Washington Press, Seattle.

Billing, J., and P. W. Sherman. (1998). Antimicrobial functions of spices: Why some like it hot. *Quart. Rev. Biol.* **73**, 3–49.

Dennett, D. C. (1995). *Darwin's Dangerous Idea. Evolution and the Meaning of Life.* Simon & Schuster, New York.

Heerwagen, J. H., and G. H. Orians. (1993). Humans, habitats, and aesthetics. In *The Biophilia Hypothesis* (S. R. Kellert and E. O. Wilson, eds.), pp. 138–172. Island Press, Washington, D.C.

Møller, A. P., and J. P. Swaddle. (1997). *Asymmetry, Developmental Stability, and Evolution.* Oxford University Press, New York.

Öhman, A., and J. J. F. Soares. (1994). Unconscious anxiety: Phobic responses to masked stimuli. *J. Abnormal Psychol.* **103**, 231–240.

Orians, G. H. (1986). An ecological and evolutionary approach to landscape aesthetics. *In Landscape Meanings and Values* (E. C. Penning-Rowsell and D. Lowenthal, eds.). Allen & Unwin, London.

Orians, G. H. (1998). Human behavioral ecology: 140 years without Darwin is too long. *Bull. Ecol. Soc. Amer.* **79**, 15–28.

Rozin, P. (1996). Towards a psychology of food and eating: From motivation to module to model to marker, morality, meaning, and metaphor. *Current Directions in Psychol. Sci.* **5**, 18–24.

Rozin, P., and A. Fallon. (1981). A perspective on disgust. *Psychol. Rev.* **94**, 23–41.

Schama, S. (1995). *Landscape and Memory.* Viking Books, New York.

Schroeder, H. W. (1989). Environment, behavior, and design research on urban forests. In *Advances in Environment, Behavior, and Design* (E. H. Zube and G. T. Moore, eds.), Vol. 2. Plenum, Press, New York.

Seligman, M. E. P. (1970). On the generality of the laws of learning. *Psychol. Rev.* **77**, 406–418.

Sommer, R., and J. Summit. (1996). Cross-national rankings of tree shape. *Ecol. Psychol.* **8**, 327–341.

Ulrich, R. S. (1993). Biophilia, biophobia, and natural landscapes. In *The Biophilia Hypothesis* (S. R. Kellert and E. O. Wilson, eds.), pp. 73–137. Island Press, Washington, D.C.

Ulrich, R. S. (1995). Effects of healthcare interior design on wellness: Theory and recent scientific research. In *Innovations in Healthcare Design* (S. O. Marberry, ed.), pp. 88–104. Van Nostrand–Reinhold, New York.

Wilson, E. O. (1984). *Biophilia.* Harvard University Press, Cambridge, Massachusetts.

AFRICA, ECOSYSTEMS OF

J. M. Lock
Royal Botanic Gardens, Kew

GLOSSARY

fynbos Habitat type in southern Africa that is characterized by thickets and low shrubs, in which fire plays a dominant role in ecosystem maintenance. Plant endemism is particularly high in these areas.

miombo Woodland habitat type widespread in south-central Africa, characterized by numerous species of the tree genus *Brachystegia* and *Isoberlinia,* which form nearly closed canopies. Fire is an annual event in this habitat, which supports relatively low populations of large mammals.

phytochorion (pl. phytochoria) Region within which a substantial proportion of the flora is endemic.

THE HUGE CONTINENT OF AFRICA STRADDLES THE EQUATOR, extending to 37°N and 35°S. It has no marginal oceanic trenches and subduction zones and so lacks the extensive mountain ranges of the Americas, but much of the southern half of the continent is a high plateau close to 1000 meters above sea level, broken only by the southern extension of the Great Rift Valley and the somewhat lower basin of the Congo (Zaïre) River. Its great latitudinal range gives it an enormous variety of climates, and this variability is reflected in an extreme diversity of ecosystems. There are three major climatic zones: two at the extreme north and south of the continent, where the main rainfall season is the winter, and one in central Africa, where the rains fall mainly in the hot summer season. The two winter rainfall areas are distinct from each other and from the tropical parts of Africa. These regions are briefly considered here; a fuller treatment appears in the entry for Mediterranean Ecosystems.

I. INTRODUCTION

White (1983) classified the vegetation of Africa into a number of phytochoria—regions within which a substantial proportion of the plants are endemic. These regions are also useful in helping to define zoogeographical regions, and comparisons can also be made with neighboring continents. Thus the Mediterranean phytochorion is most closely related floristically to southern Europe and the Middle East. The Somali–Maasai phytochorion, which occupies the Horn of Africa and regions south to central Tanzania, shows floristic relationships with parts of Arabia and the western part of the Indian subcontinent. On the other hand, about 80% of the plant species that occur in the Guineo–Congolian phytochorion, which occupies the Congo Basin and extends westwards to Liberia, are endemic. The Cape Region is also extremely rich in species, almost all of which occur nowhere else.

The large island of Madagascar, off the southeast coast of Africa, has been isolated from the rest of the continent for over 100 million years and its flora and fauna are very different indeed from those of Africa. The ecology of Madagascar is treated, in brief outline, in Section V.

II. MAJOR ENVIRONMENTAL FACTORS OF CONTINENTAL AFRICA

A. Geology

Much of southern tropical Africa is a plateau lying at or near 1000 m above sea level, most of which has not been submerged since the Tertiary period. This plateau is formed of Pre-cambrian ("basement complex") rocks, mainly igneous and much metamorphosed, which have been heavily and deeply weathered. The topography of much of the plateau is gently rolling, often with isolated rocky inselbergs, and sometimes with flat-topped hills capped with secondary ironstone, a product of prolonged weathering in high temperatures.

In eastern Africa the plateau has been cut by the two branches of the Great Rift Valley, which extends from Israel to the Zambezi River. Rifting seems to have begun in the Miocene, and was (and still is) accompanied by volcanic activity. In Ethiopia great sheets of basalt were erupted and now form the Ethiopian Highlands. Scattered volcanoes line the branches of the Rift Valley; some, such as Elgon and Kilimanjaro, are now extinct or dormant, and others, such as O1 Doinyo Lengai (Tanzania) and Nyiragongo (Zaïre), are still active. The Rift Valley contains numerous lakes, some relatively shallow and often saline [e.g., Lake Chala (Ethiopia), Lakes Turkana and Bogoria (Kenya), and Lakes Natron and Eyasi (Tanzania)] and others deep and freshwater [Lakes Albert and Edward (Uganda/Zaïre), Lake Tanganyika, and Lake Malawi]. The blocks between the branches of the Rift Valley have tilted as part of the same tectonic disturbances. This has impeded or even reversed the flow of rivers, and led to the formation of extensive shallow lakes, such as Lake Victoria, and huge areas of swamp.

B. Climate

The climate of Africa (excluding the extreme north and south) is determined by the movement of the Inter-Tropical Convergence Zone (ITCZ). As the apparent position of the sun moves north and south with the seasons, the area where the sun is immediately overhead is heated more than the areas to the north and south. The heated air rises, and air is drawn in from north and south to replace it. The convergence of these two air masses, and the rising of the heated air, causes rain. The rain belt tends to lag behind the sun, so the rain belt reaches its farthest north in July–August, rather than at the summer solstice (June 21st), and its farthest south in January–February, not at the winter solstice (December 21st). Away from the equator, the rainfall produced by the ITCZ as it moves in one direction merges into that produced during its return, so that there is a single rainy season. Near the equator, however, there is a tendency for the two passages of the ITCZ to be separated by a dry season, giving two rainy seasons and two dry seasons in each year; in the wettest regions these dry gaps hardly occur so that rainfall occurs more or less throughout the year.

Superimposed on this basic seasonal pattern is variation in the total annual rainfall. As a rule, totals decline away from the equator, and also from west to east. In East Africa, air coming from the north-east has passed over the dry Arabian land-mass, not the sea as in most of the rest of the continent, and this explains the lower rainfall. Areas close to the sea tend to be wetter than inland areas, except where, as in south-western Africa, cold sea currents offshore produce foggy conditions but low rainfall. Extremes include Sierra Leone, where the coastal capital, Freetown, has over 250 cm of rain each year, most of it falling in four months. Debunscha Point, on the seaward margin of Mt. Cameroon, receives over 10 m of rain in some years; this comes from the combination of a warm sea and rapidly rising ground close to the shore. Altitude also affects rainfall; totals appear to increase up to about 3000 m and then decline; the upper regions of Mt. Kilimanjaro are a virtual desert with probably less than 25 cm of rain each year. At middle and higher altitudes, the rain is supplemented by clouds from which droplets of moisture condense onto leaves and drip to the ground, and here humidity is extremely high for much of each day.

C. Soils

The soils of much of Africa are developed on very ancient land surfaces and are therefore the product of long weathering. Forest soils, developed under high rainfall, are generally very strongly leached, rich in clay, and yellowish brown in color. Drier forests have redder, somewhat less nutrient-poor soils. Savanna soils are often brownish and, again, rich in clay. On the older plateaux there are often regular sequences of soils ("catenas"), with leached sandy or gravelly soils with kaolinite clay minerals on the hill and ridge tops, finer-grained soils on the hill slopes, and deep, dark-colored clays in the valleys. These valley soils, which shrink and crack as they dry and swell again when they are wetted, become extremely sticky and difficult to work when they are seasonally flooded or waterlogged. These are the so-called "cotton-soils," rich in montmorillonite clay minerals, in which motor vehicles so easily become bogged.

As a result of prolonged weathering, most of the soils of the African plateau are nutrient-poor, with phosphorus in particular often being in short supply. Some of the more volatile elements, such as nitrogen and sulfur, can be lost in the smoke in the regular dry season fires.

In a few areas, such as the "copperbelt" of southern Zaïre, northern Zambia, and the Great Dyke in Zimbabwe, there are outcrops of soils rich in metals such as copper, cobalt, and chromium. These produce soils that are toxic to many plants. Often trees are scarce or absent, and the vegetation is mainly composed of grasses and a range of specialized herbs, often very local in their distribution.

D. Fire

The highly seasonal climates of most of tropical Africa provide ideal conditions for widespread vegetation fires. During the wet season, grass growth is rapid, but during the dry season the aerial parts of the grasses dry out and burn very easily. Most of those parts of Africa with a seasonal climate are burned by vegetation fires every year. The exceptions are areas with a rainfall too low to produce the necessary volume of fuel, or so high that the tree cover is continuous and dense, preventing the growth of enough grass to provide fuel. The carbon dioxide produced by this biomass burning is highly significant in the annual atmospheric carbon dioxide budget of the world. There can be little doubt that fire has been important in African vegetation patterns for a very long time. Natural fires, at long intervals, can be started by lightning, as well as by volcanic activity. Once humans began to use fire—and some have suggested that this may have been as much as a million years ago—fire frequency would have increased. The acquisition of the ability to make (not just use) fire would have increased this frequency further, and another increase surely came when safety matches became widely available.

The question of the influence of fire on the distribution of vegetation types in Africa has been extensively discussed. On the one hand there are the views of workers such as A. Aubréville, who considered fire to be a major destructive force that had probably been responsible for the destruction of vast areas of the drier forests of Africa. The converse view, that fire has been a feature of African vegetation for so long that it is now a fundamental part of many ecosystems, is probably more widely held. Though fire may sharpen boundaries between vegetation types, it does not greatly affect their distribution. Assessment of the importance of fire also depends on viewpoint. Foresters generally disapprove of fire and expend much effort in its prevention, whereas pastoralists use fire extensively to remove old grass and to stimulate the development of young growth, as well as to check development of woody plants. Foresters, who want trees, generally recommend that fires should be started early in the dry season, when the fuel is still not completely dry ("early burning"). Pastoralists, who want grass, prefer a fire late in the dry season when the fuel is completely dry and damage to woody growth is maximal ("late burning").

The presence of specific adaptations to fire in a number of plants is a further argument for fire having been an important factor for a long time. Several trees, such as the shea-butter tree (*Vitellaria paradoxa*) and species of *Combretum* and *Pterocarpus*, have seeds that, when they germinate, produce what appears to be a radicle that pushes into the ground. However, close examination shows that at least part of this is hollow and is actually formed by fusion of the stalks of the cotyledons (seed-leaves), which remain within the seed. The terminal bud, which lies between the bases of these stalks, is thus carried down below ground level so that it is at least partially protected from fire. Once well buried, the bud starts growth and the shoot breaks out through the side of the apparent root. It elongates and eventually emerges from the ground, but not before producing a number of reduced scale-like leaves, each with a bud in its axil. If the terminal bud is burned away, the lateral buds in the axils of the scale leaves start to develop, giving the seedling another chance of establishment.

Many of the grasses of the fire-swept regions have a long awn on their grains. This awn is hygroscopic, twisting and untwisting with daily changes in humidity. The tip of the grain is sharply pointed and bears stiff hairs. If the point enters a crack, the hairs prevent it from coming out again, and the regular movements of the awn will tend to drive it deeper into the soil, until it is protected from the main heat of the fire.

E. Past Climatic and Environmental Fluctuations

Tropical Africa has not escaped the climatic fluctuations of the last 2 million years. However, interpretation of changes in Africa is complicated by the effects of rifting and vulcanism, particularly in the east, as well as by the presence of many isolated mountains and mountain ranges. Evidence is also relatively scarce because the peat and other organic deposits that are widespread in cooler regions do not accumulate nearly so widely in tropical lowlands because high temperatures speed decomposition processes. The peat deposits that do exist are mainly on mountains, and so may not present a representative picture of the much more extensive lowlands.

However, there are other kinds of evidence available, such as the extent of active and inactive (fossil) sand dunes. In West Africa the fossil dunes extend 400–600 km south of the present limit of active dunes; the arid period during which these fossil dunes formed appears to have been 12,000 to 20,000 years ago. South of the equator, the Kalahari Sands extend far to the north and east of any presently active dunes, reaching the Zaïre River in the north. It appears that the most recent period of active dune growth coincided with that north of the equator. Lake levels, as marked by raised beaches, can also provide evidence, at least of times when lake levels stood higher than they do now. The study of plant distributions can also sometimes suggest former connections between similar but now isolated vegetation types. For instance, several tree genera with most of their species in West Africa have one or more species in the Eastern Arc montane forests of Tanzania. This suggests a connection between these forests areas in the distant past.

In general, periods marked by cold in high latitudes are correlated with periods of drier and cooler climate in tropical Africa. During these periods, forested regions contracted and "savanna" regions spread. In the driest times, forests seem to have contracted into a number of refugia, in Sierra Leone and Liberia in West Africa, as well as Cameroon and Gabon, eastern Zaïre, and eastern Tanzania. It is now agreed that most of the evidence shows that there was an arid period between 12,000 and 20,000 years ago. Prior to this there were certainly similar fluctuations, perhaps of greater magnitude, but their dating and extent are still somewhat in dispute. Since 12,000 years ago there have also been changes; the wettest period seems to have been between 8000 and 4000 years ago. At this time, Neolithic pastoralists inhabited large areas of the Sahara and left rock paintings showing elephants, giraffes, and antelopes in areas now far too dry to support them.

III. MAJOR PHYTOGEOGRAPHIC AND ECOCLIMATIC ZONES OF CONTINENTAL AFRICA

A. Mediterranean North Africa

This region borders the Mediterranean Sea, with winter rainfall and hot dry summers. It includes Morocco, Algeria, Tunisia, Libya, and Egypt. The wettest part is the west, and here also there are the high mountains of the Atlas range. Thousands of years of human settlement, agriculture, and grazing of domestic animals have greatly altered the ecosystems. In Roman times there were certainly lions (*Panthera leo*) (and therefore a substantial prey population) and probably elephants (*Loxodonta africana*) in this region, but all are now gone.

The wetter parts were probably originally covered with forest, but this is now represented only by tiny fragments; *Celtis australis* and *Pistacia atlantica* may have been important trees in the original forests. The drier forests were (and in places still are) dominated by evergreen oak (*Quercus ilex*), which casts a dense shade in which few other species can grow, or by cork oak (*Quercus suber*). There are also coniferous forests of species such as Aleppo pine (*Pinus halepensis*) and North African cedar (*Cedrus atlantica*). Other parts are covered by scrub, similar in physiognomy to the chaparral of California and the fynbos of the Cape Region of South Africa, made up of shrubs with small hard (sclerophyllous) leaves, such as the kermes oak (*Quercus coccifera*), wild olive (*Olea europaea*), and, in very degraded sites, the dwarf palm *Chamaerops humilis*. The shrubs are fire-resistant, sprouting from the base after fires. The gaps between the cushions and, lower down, between the sclerophyllous shrubs support a rich herb flora including many annuals (Fabaceae are abundant and diverse) and many plants springing from underground bulbs or corms (geophytes). These grow during the late winter and spring, flower, and then dry up in the baking heat and drought of the summer. At high altitudes, between 2800 and 3800 m, there is often a low scrub made up of spiny cushions, sometimes graphically referred to as hedgehog heath.

B. The Cape Region of South Africa

The Cape Region also has a Mediterranean climate, although, being in the Southern Hemisphere, it enjoys a hot dry summer when North Africa is having a cool wet winter. Like North Africa, thicket and low scrub (known locally as fynbos) are the main physiognomic vegetation types, and fire is a regular influence on the vegetation. Many species appear to be fire-adapted. Some species, such as the red-flowered, lily-like species of *Cyrtanthus*, flower only after fires, stimulated either by chemicals in the smoke or by the greater daily fluctuations in soil temperature that follow removal of the vegetation cover. Others, such as species of *Leucadendron* (a *Protea* relative), retain their seeds on the parent plant and only release them after fire. The range of fire-adapted species in fynbos, and the many ways in which they respond to fires, suggests that fire has been a feature of this vegetation type for a very long time.

The region is extremely diverse geologically and has numerous isolated mountain ranges. These factors, combined with long isolation, have given rise to an extraordinary diversity of plant species—estimates vary from about 7000 species in the 71,000 km^2 of the region (White, 1983) to 8600 species in an area of 91,000 km^2 (Cowling and Richardson, 1995). The genus *Erica* (heaths) has over 500 species in the Cape Region; other extremely diverse genera include *Aspalathus* (Fabaceae, 250 species) and *Muraltia* (Polygalaceae, 100 species). The family Proteaceae, including 85 species of *Protea*, is prominent among the larger woody plants, and is associated with an endemic pollinator, the Cape Sugarbird (*Promerops cafer*). The lower-lying and more level areas have largely been converted to agricultural land, but the mountain ranges continue to provide refuges for the endemic flora, although invasion by woody species introduced from other regions with a similar climate is a major problem.

Originally there was a rich fauna of large mammals but these were heavily hunted by European settlers. Some still survive in reserves, but the quagga (a form of zebra, *Equus quagga*) is extinct, and the bontebok (*Damaliscus dorcas*) and white-tailed gnu (*Connochaetes gnou*) survive only on enclosed farms.

C. Tropical Africa

The region known as tropical Africa takes in most of the continent. The vegetation is determined by the climate, which is highly seasonal over much of the region. There are virtually no parts of Africa without some kind of a dry period; truly ever-wet climates like that of Singapore (where two weeks without rain is a drought) are virtually absent. The major exception lies along the equator in West Africa; here the dry season lasts a month or less. The equatorial regions of eastern Africa, however, lie within the rain-shadow of the Arabian landmass, and here, even on the equator, rainfall is low. Furthermore, there tend to be two rainy seasons rather than one. Both of these vary in intensity, and the gaps between them also vary in length, so that this region tends to suffer more than most from periodic droughts.

Tropical Africa can be divided into two parts. South and east of a line from Ethiopia to the mouth of the Zaïre River, most of the land forms a dissected plateau lying at about 1000 m above the sea. This is split from north to south by the Great Rift Valley, which extends from Israel through the Red Sea, then across Ethiopia, Kenya, and into Tanzania. A western branch runs south through Uganda, along the western side of Tanzania, and ends in Malawi. Between the branches of the Rift Valley, the land surface has tilted in places, disrupting river flows and producing the huge but shallow Lake Victoria as well as the extensive swamps of Uganda. To the north and west of the high plateau, the general land surface is much lower. Here again rocks of the basement complex underlie most of the region, but younger rocks are found here and there. Most of tropical Africa is covered by woodland and various forms of savanna, with forest occupying the basin of the Zaire River, and drier bushlands, thickets, and grasslands in the equatorial regions of eastern Africa.

IV. MAJOR ECOSYSTEMS OF TROPICAL AFRICA

A. Forest

True tropical forest is confined in Africa to two main blocks: the basin of the Zaïre (Congo) River, extending north and west through Cameroon and Gabon into southern Nigeria and east to the borders of Uganda, and a region farther to the west stretching from western Ghana through Ivory Coast and Liberia into Sierra Leone. Elsewhere, isolated forest patches are found in and around the Ethiopian Highlands, and along the East African coast. Most forests are found where annual rainfall exceeds 120 cm and where the dry season is no more than four months long. At the margins, small differences in water availability can greatly affect the vegetation; forest may extend far into grassland along the banks of rivers as "gallery forest," and non-forest plants can be found well within the forest zone on rocky outcrops.

African forests are poor in species compared to those of South-east Asia and the Amazon Basin, but they are still species-rich places. Surveys in Ghana found forests in which a 25-by-25-m patch contained more than 200 higher plant species. These surveys also showed that the wettest forests are the most species rich. They do not, however, contain the tallest trees, which are found in drier, partially deciduous forest, perhaps because the soils there are not so leached of nutrients as those in the wetter areas. These drier forests are, at least in Ghana, richest in the number of timber trees. The most prominent plant families in the forest canopy are the mahoganies (Meliaceae) and legumes (Fabaceae, particularly the subfamily Caesalpinioideae).

Following disturbance, the pioneers such as umbrella trees (*Musanga*, in Moraceae) and *Trema* (Ulmaceae) grow at 3–4 m per year and rapidly form a tree cover. These species shade the soil and improve conditions for the establishment of trees whose seeds can germinate in shade, and whose seedlings can tolerate low light intensities. These early-successional species include some of the mahoganies, such as *Khaya* and *Entandrophragma* (sapele), whose seeds are wind dispersed. Studies in Uganda have suggested that eventually this "mixed forest" may give way to a more species-poor forest in which leguminous trees are common. Some of these leguminous trees form extensive areas of forest in which a single tree species is dominant—an unusual state in tropical forest for which there is currently no satisfactory explanation. *Cynometra alexandri* in eastern Zaire and Uganda and *Gilbertiodendron dewevrei* in the Zaire River Basin form monospecific forest stands.

Beneath the canopy trees grow smaller trees and shrubs, some of them young plants of canopy species, but others naturally small at maturity. It is usually quite easy to walk through undisturbed forest, as there are few herbs on the forest floor. Only where a tree has fallen, allowing more light to reach the ground, is there a dense mass of quick-growing herbs in the ginger (Zingiberaceae) and arrowroot (Marantaceae) families. Woody climbers (lianas) are also most frequent in disturbed forest, and can reach the tops of the tallest trees.

African forests are rich in animal species. The canopy is occupied by many species of monkey. Some, like the species of colobus, are leaf-eaters. Others, such as the various small guenons (*Cercopithecus*), feed mainly on fruit. Different species of guenon specialize on different dietary mixes and several species can thus coexist in the same area of forest. Other tree dwellers include species of flying squirrel, and also true squirrels, most of them seed-eaters.

The great apes—gorilla (*Gorilla gorilla*), chimpanzee (*Pan troglodytes*), and bonobo (*Pan paniscus*)—travel mainly on the ground and feed both there and in the trees. All live in groups and exploit extensive home ranges that they know intimately, moving from one seasonal food resource to another. A high proportion of their food is fruit, and they are important dispersers of seeds. In the wettest and least seasonal forests there are also large ground-dwelling monkeys—the drill (*Mandrillus leucophaeus*) and the mandrill (*Mandrillus sphinx*). Like the apes, they are highly social, living in troops of a hundred or more individuals and traveling over a huge home range.

There are also ungulates in the rain forests of Africa, ranging in size from the tiny royal and dwarf antelopes (*Neotragus pygmaeus* and *N. batesii*, 25–30 cm at the shoulder) and the taxonomically very distinct water chevrotain (*Hyemoschus aquaticus*, only 30–40 cm), through several species of duiker (*Cephalohus*, small antelopes), to the large bongo (*Tragelaphus euryceros*). The Ituri forests of eastern Zaïre harbor, the okapi (*Okapia johnstoni*), whose closest relative is the giraffe. All of these are browsers on the leaves of forest shrubs and herbs.

The largest forest animal is the elephant. Forest elephants are usually recognized as belonging to a separate subspecies from those of open country; they are smaller and tend to have smaller and straighter tusks. Elephants are voracious feeders; their inefficient digestive systems mean that they must feed for a large proportion of each day to obtain enough nourishment. They are particularly fond of clearings in forest, because here more of the foliage is close to the ground and accessible, and by concentrating their feeding in such areas they may prevent tree development and perpetuate the clearings. Elephants also eat large quantities of fruit. Plants such as *Panda oleosa, Desplatsia*, and *Balanites wilsoniana* have such large seeds that it is hard to know what animal other than an elephant could possibly disperse them. Attempts to germinate seeds of *Panda* that have not passed through an elephant have been unsuccessful. The effects on such plants of the widespread decline in elephants may be severe.

B. Seasonal Tropical Vegetation

Much of tropical Africa is occupied by vegetation that develops under a climate in which the year is divided into dry and wet seasons. The vegetation of these areas is of many kinds, ranging from woodland with an almost closed canopy at one extreme to dry open grasslands at the other. All of these vegetation types have been referred to as savanna, but recently there has been a tendency to attempt to discard this term. It has been argued that a word that can mean almost anything from closed forest to open grassland is too vague to have any utility in ecological discussions. White's classification avoids the term, and instead uses "woodland," "wooded grassland," "grassland," "bushland and thicket," and "shrubland" to define the various physiognomic vegetation types more precisely. These terms will be used here, but where the term "savanna" has wide currency, as in much of West Africa, it will be used.

In tropical seasonal vegetation, water is abundant and growth can be rapid during the wet season, but in the dry season the grass rapidly dries out and becomes flammable; fires are frequent and often annual. It is somewhat paradoxical that the fiercest fires occur where the rainfall is highest—and where, therefore, biomass production during the growing season is highest. Over much of the savanna regions of Africa there is one wet season and one dry season each year, but near the equator, particularly in East Africa, there can be two wet and two dry seasons each year (see Section II,B), so that the seasonal cycle is six months rather than twelve.

The herds of wild ungulates and their associated predators for which Africa is so famous are all found in areas of seasonal tropical vegetation. Densities vary enormously, now largely because of human pressures, but there can be no doubt that densities varied greatly before human impacts became significant.

1. Woodland

In West Africa, where the isohyets (lines of equal rainfall) are more or less parallel to the coast, there are belts of vegetation that follow the isohyets. These were classified many years ago by the great French botanist Auguste Chevalier into three savanna types: Guinea, Sudan, and Sahel. In West Africa, use of the term "savanna" is long established, and Chevalier's zones are still recognized and used today. The southernmost zone, with the highest rainfall and the shortest dry season, is the Guinea Zone.

Woodland has been called by many different names in the African ecological literature. Tree savanna is perhaps the most widely used, as well as the more descriptive "tall grass–low tree savanna." In West Africa, woodland is referred to as "Guinea Savanna," and in francophone countries it is often called "forêt claire." This is a confusing term, because it implies that woodland is essentially the same as true forest ("forêt dense"). This is not the case; surveys in Nigeria of adjacent forest and woodland have shown that they have very few species in common. Physiognomically, woodland consists of a single tree layer, sometimes with an almost closed canopy. Lianas are absent, as are epiphytes. Grass covers the ground beneath the trees.

The trees are often misshapen because the frequent fires kill the growing points and cause branching. Because grass productivity is high, fires are more intense than in other zones. The bark of many of the trees is thick and furrowed, which insulates the delicate growing tissue (cambium) from the heat of fires. Regeneration is difficult as young seedlings are in the hottest part of the fire. Some species, such as the shea-butter tree (*Vitellaria paradoxa*), have specialized germination mechanisms that may help to overcome this. Some grass seeds, mainly those from the subfamily Andropogoneae, also have a mechanism for evading fires (see earlier discussion).

The greatest obstacle to the development of the woodlands of Africa (many would say their most valuable conservator) is the presence of the tsetse fly (*Glossina morsitans, G. pallidipes*, and other species). This biting fly is the vector of trypanosomiasis, a lethal protozoan disease of cattle and of sleeping sickness in humans. Wild game animals are immune to trypanosomiasis, although they can carry the protozoan in their blood. Control of the fly has been attempted in many areas, using methods of varying destructiveness to the ecosystem. The fly requires shaded places in which to rest, so tree clearance has been used. Aerial spraying with DDT and other non-specific insecticides has now been replaced by highly specific treatment with persistent insecticides of the undersides of large tree branches, where the flies rest. Other attempts to break the cycle of transmission have included the shooting of wild ungulates. Researchers continue to look for effective and environmentally sound ways of tsetse control. An alternative that has been tried in some areas is to leave the flies alone and harvest the wild game instead of the cattle. This has an advantage in that each member of the diverse ungulate community occupies a different niche, so that the carrying capacity and yield are, at least in theory, higher than those from a monoculture of cattle.

The Guinea Savanna woodlands of West Africa have as their characteristic tree species *Lophira alata* (Ochnaceae) and *Daniellia oliveri* (Fabaceae). Slightly drier woodlands have species of *Isoberlinia* (Fabaceae). The main grasses are species of *Andropogon* and *Hyparrhenia*, both with awns on their grains. These large grasses stand 2–3 m high when mature. Many agronomists have been misled into thinking that this high standing crop also implies a high animal production potential, but this is not so. When the grasses are at their tallest, they are made up largely of stem, which is hard, woody, and of very low nutritional value. After a fire, and at the beginning of the rains, nutritious foliage is produced, but this does not last long.

This extreme seasonality in the quality and availability of grass may go a long way towards explaining the relative scarcity of large wild animals in this zone. Various antelopes, such as roan (*Hippotragus equinus*) and eland (*Taurotragus oryx*), occur here, as do smaller species such as bushbuck (*Tragelaphus scriptus*). Elephants occur at least seasonally. Predators are correspondingly also scarce, with leopard (*Panthera pardus*) and lion both occurring at very low densities.

The corresponding zone to the Guinea Savanna south of the equator is the miombo woodland. "Miombo" is the local name for one of the commonest trees, *Isoberlinia tomentosa* (Fabaceae), but by far the most widespread and abundant trees belong to the genus *Brachystegia* (Fabaceae). There are about 25 species of this genus in the woodland regions of south-central Africa, all fairly similar to one another, mostly variable, and difficult to identify. Miombo covers vast areas of southern Tanzania, Zambia, northern Zimbabwe, southern Zaïre and south-eastern Angola. Miombo is physiognomically rather different from Guinea Savanna; the trees are usually taller, often with an almost closed canopy, and the grass is somewhat shorter. However, fires are virtually an annual event. Although at first sight miombo woodland can appear rather monotonous, over much of its extent it shows considerable local variations related to topography. The ancient land surface on which it grows is weathered into a pattern of gentle hills and valleys.

The tops of the hills and ridges are dry and well drained, but there are often rock outcrops (inselbergs) around which water accumulates so that the vegetation around them is tall and dense. The rocks themselves, if not sloping too steeply, often bear a shallow turf that dries out completely during the dry season, but during the wet season it becomes saturated and supports many small ephemeral plants such as species of *Utricularia* (bladderworts), *Xyris* (yellow-eyed grasses), and various sedges (Cyperaceae). There are often mats of the sedge *Afrotrilepis*, which can dry out completely and then rehydrate and resume growth when water is once again available. In southern Africa there can be clumps of the shrub *Myrothamnus flabellifolius*, the resurrection plant, which behaves in the same way (i.e., it is poikilohydric). These rocky hills are often home to a small and specialized antelope, the klipspringer (*Oreotragus oreotragus*), and to rock hyraxes (*Procavia johnstoni* and *Heterohyrax brucei*).

The typical miombo woodland occurs on the hill slopes, but even this is not uniform. Termite mounds, built by species of *Macrotermes*, can be up to 4 m high and 10 m in diameter. These termites are fungus gardeners; they collect dead plant material from a wide area and carry it into the center of the nest. Here it is formed into a honeycomb-like structure that is colonized by fungi that can break down the cellulose and lignin in the wood. The termites feed on the fungal hyphae and are thus indirectly nourished by the wood. A by-product of this activity is the concentration of mineral nutrients and fine soil fractions in the mound so that, somewhat paradoxically, it becomes a favorable habitat for plant growth. Termite mounds bear much denser woody vegetation, often casting enough shade to exclude grasses and therefore fires.

The valley bottoms in the miombo ecosystem are seasonally waterlogged, and this prevents the growth of most trees. Tall grassland occupies this habitat, the grasses often forming well-spaced tussocks between which small plants can grow early in the wet season. The soils of these valley grasslands are often black clays that are extremely sticky when wet and very hard when dry.

As in the Guinea Savanna, large mammal biomass is low in miombo. There are, however, many more species than in Guinea Savanna. Roan antelope occur, as well as the closely related sable antelope (*Hippotragus niger*) with its spectacular scimitar-shaped horns. Hartebeest (*Alcelaphus busephalus*) occur as do elephant where there is permanent water within reasonable range. The spectacular greater kudu (*Tragelaphus strepsiceros*), with long, spirally twisted horns, is very much an animal of the miombo. Herds of Cape buffalo (*Syncerus caffer*) can be found, again where permanent water lies within a reasonable distance.

The annual cycle of the miombo woodland follows the rains. The trees often begin to produce their new leaves before the first rains, presumably drawing on reserves of water from deep in the soil. This flush of new foliage is often reddish, perhaps because of secondary compounds that make the leaves unpalatable to predators. Flowering takes places fairly early in the wet season, giving the often bulky pods and seeds time to develop before the dry season begins. The pods ripen and burst to scatter their seeds during the dry season. The extent to which the leaves are shed varies from place to place and year to year according to the degree of drought and the intensity of the fire.

Drier woodlands are found both to the north and south of the equator. To the north, they fall into the zone called "Sudan Savanna" by Chevalier. This zone has a longer dry season and a lower annual rainfall than the Guinea Savanna. Typical tree species include *Parkia filicoidea* and *Piliostigma thonningii*, but this zone has a much higher human population and is more cultivated than the Guinea Savanna. There are a number of possible explanations for this; maybe all play a part. Because it is farther from the coast, it was less heavily raided and depopulated during the time of the slave trade. The relatively smaller growth of grass makes land clearance and maintenance easier, and the lower rainfall means that the soils are less intensively leached and therefore rather more fertile. Whatever the reason, this zone now supports high human populations who cultivate groundnuts, sorghum, and various kinds of millet. Many of the remaining trees survive because they are of some use; *Parkia* seeds and the pulp that surrounds them are a useful food, as are the leaves and fruit-pulp of the baobab tree (*Adansonia digitata*).

In southern Africa the corresponding zone is probably mopane woodland, which is not very intensively settled or cultivated. It is dominated by a single tree species, *Colophospermum mopane* (Fabaceae: Caesalpinioideae). Mopane grows in areas that are hotter and drier than those occupied by miombo. It is unusual in its family in being normally wind pollinated. Mopane woodland can support a wide range of large mammals, including elephant, black rhino (*Diceros bicornis*), eland, and impala (*Aepyceros melampus*).

2. Bushland and Thicket

The drier lowlands of eastern Africa and the Horn of Africa are occupied by woody vegetation that is low-growing (less than 10 m tall), often spiny, and leafless for a substantial part of the year. The rainfall in this region is always somewhat unpredictable, tending to fall in two wet seasons in each year, either or both of which may fail. The most obvious components of the vegetation are species of *Acacia* (Fabaceae: Mimosoideae). Almost all of these bear paired spines, sometimes hooked and very sharp. Another genus that is abundant and diverse here is *Commiphora* (Burseraceae); all the trees in the genus have a resinous scent and one is the main source of myrrh. [Another genus of the region, *Boswellia* (Burseraceae), yields the resin known as frankincense.]

Many other tree and shrub species also occur so that this vegetation type is very rich in species. Few of these are common to the wetter woodland regions, and the Somali–Maasai area is recognized by White as a separate phytochorion, that is, a region with a distinctive flora. The floristic relationships of this area lie with southern Arabia and the north-western part of the Indian peninsula rather than with the rest of Africa. Somalia and northern Kenya in particular are rich in endemic plant species, perhaps partly because of long isolation, but also because of the existence of specialized habitats such as regions where the underlying rock is either limestone or gypsum (calcium sulfate) and the coastal strip near Obbia (Somalia), where low-growing plants cover fossil sand dunes just inland from the coast.

The animals of this area are also distinctive. Grevy's zebra (*Equus grevyi*) and the wild ass (*Equus africanus*) are virtually confined to this part of Africa, and the latter is very scarce. The very handsome and distinct reticulated giraffe (*Giraffa camelopardalis reticulata*) is confined to Somalia, southern Ethiopia, and northern Kenya. The beira (*Dorcatragus megalotis*), a specialized long-legged duiker, is confined to Somalia. The four species of dikdik (*Madoqua*), tiny antelopes that live alone or in pairs in bushland, in territories marked by dung piles, are virtually confined to this vegetation type. Several species of gazelle are also confined, or almost so, to this region. Finally, two remarkable long-necked gazelles, the dibatag (*Ammodorcas clarkei*) and the gerenuk (*Lithocranus walleri*) occur only here; both are exclusively browsers, stretching their long necks up into the bushes and often standing on their hind legs to reach taller growth. Their narrow muzzles allow them to pick out small leaves from between dense twigs or thorns.

The islands of Socotra and Abd-el-Kuri lie to the east of the tip of the Horn of Africa. Both have undoubtedly been isolated for a very long time and have no indigenous large mammals (although introduced livestock now abound). This has allowed the development of a remarkable flora that, although clearly part of the Somalia–Maasai Region, includes numerous endemics such as *Dendrosicyos socotranus* (a tree-forming representative of the cucumber family), *Dorstenia gigas* (Moraceae), and *Adenium socotranum* (Apocynaceae).

3. Grasslands

Many of the grasslands of tropical Africa are the product of some special condition of the soil that prevents the growth of trees; seasonal waterlogging, shallow soils subject to extreme seasonal droughts, and high concentrations of metals such as copper and cobalt all lead to local grasslands. However, near the equator in eastern Africa there are extensive areas of grassland, sometimes with scattered flat-topped acacia trees (*Acacia tortilis*) or thicket clumps. Most of these areas lie between 1000 and 2000 m above sea level. The commonest and most prominent grass is the red oat grass (*Themeda triandra*). This species thrives under a regime of annual burning and light grazing, but it is vulnerable to overgrazing. Dense tussock-forming grasses such as *Sporobolus pyramidalis* tend to replace it if the grazing pressure is excessive.

These grasslands have long been the territory of pastoral peoples such as the Maasai, and it is possible that these people, who are well aware of the effects of fire on vegetation, have used it over the millennia to alter the balance from woodland or thicket to grassland on which to pasture their cattle. These peoples have also long lived in close proximity to huge populations of wild ungulates and their predators. These ungulate populations often make seasonal migrations to make best use of their range, such as that which straddles the Tanzania–Kenya border in the Serengeti and Mara region. The main species in these migratory populations are wildebeest (*Connochaetes taurinus*) and the plains zebra (*Equus quagga boehmii*); other more sedentary species include Thomson's gazelle (*Gazella rufifrons thomsoni*) and Grant's gazelle (*Gazella granti*). In wetter areas, with access to permanent water, other sedentary species such as Cape buffalo, elephant, and waterbuck (*Kobus ellipsiprymnus*) are found. The Serengeti migration is the best-known, but in southern Sudan there is also a huge migratory system involving tiang (*Damaliscus lunatus*), Mongalla gazelle (*Gazella rufifrons albonotata*), and white-eared kob (*Kobus kob leucotis*). These species spend the wet season in the tall grasslands towards the Uganda border. When the dry season begins, they move northwards and feed on the swamp grasses, such as *Oryza* (wild rice) and *Echinochloa*, exposed by the retreat of the Nile River floods. This migration crosses the line of the partially dug Jongei Canal—now abandoned due to civil war—which forms a significant obstruction.

The biomass of ungulates supported by some of these grasslands is very high, particularly where there are two rainy seasons each year and therefore at least some forage of good quality at all times. The Western Rift Valley in Uganda and eastern Zaïre is typical; here large herds of elephant, hippopotamuses from the lakes, and Cape buffalo form the main part of the biomass, but other species such as Uganda kob (*Adenota kob thomasii*), waterbuck, and warthog (*Phacochoerus africanus*) are also common.

In recent years, following the widespread destruction of the larger animals, the smaller ones have increased in numbers. The killing of many of the elephants has also led to widespread vegetation changes. *Acacia* trees, whose seedlings were formerly so regularly browsed and burned that they remained in a suppressed state in the grass, are now regenerating in many places, and thicket clumps dominated by the thorny scrambler *Capparis tomentosa* (Capparidaceae) are now spreading and coalescing to produce larger areas of thicket that are becoming home to giant forest hogs (*Hylochoerus meinertzhageni*), a species more commonly found in montane forests.

In all of these regions with numerous herbivores, there are high populations of predators. Lions are the most conspicuous, living in prides made up of one or more females and their offspring with at least one mature male in attendance. The females collaborate in hunting, and any kill is shared by all the pride. When a new male displaces another, his first action is to kill any cubs in the pride still dependent on their mothers. This action rapidly brings the lionesses into estrus and allows the new male to start passing on his genes with the minimum of delay.

Spotted hyenas (*Crocuta crocuta*) can also be very common and have been shown to function not only as scavengers but also as highly efficient predators in their own right. They live and hunt in matriarchal groups that hold and defend communal territories. The cheetah (*Acinonyx jubatus*) is a highly specialized and solitary cat, entirely dependent on its speed for running down prey. Its jaws are relatively weak, and many of its kills are lost to lions or hyenas against which it has virtually no defense. Finally, there is the wild dog (*Lycaon pictus*). This is another species that lives in groups; hunting is by a prolonged chase in which the members of the pack take turns in the lead. The leaders snap and tear at the hind end of the prey, eventually bringing it down. Wild dogs are short-lived, however, and rely on frequent large litters to maintain their numbers. They usually have a very wide home range and in many areas they are now very scarce or absent, and may well, with the cheetah, be Africa's most endangered carnivores.

The African grasslands are also home to large ground-living birds; some, like the ostrich (*Struthio camelus*), are flightless and others, like the kori bustard (*Ardeotis kori*), almost so. The large mammal populations provide food for scavengers; for instance, a carcass on the Serengeti Plains will attract four or five species of vulture. Vultures spend much of their day soaring on thermals; they watch both the ground and each other so that if one spots prey and starts to descend, others quickly follow.

4. Shrublands

Shrubland occupies the driest areas; under conditions of even lower rainfall the individual shrubs grow farther and farther apart until the land is best referred to as desert. The shrubs are generally between 10 cm and 2 m in height, and are of many different species and families. Members of the Acanthaceae, particularly the genera *Barleria* and *Blepharis*, are often prominent. Many are spiny, and all are highly facultative and irregular in their production of leaves and in their flowering and fruiting, often not reproducing every year but only when rainfall is exceptionally high or prolonged. Under these conditions, many annual grasses also appear between the shrubs. Perennial drought-resistant grasses often occur among the shrubs, and some authorities believe that the natural state in these ecosystems is a drought-resistant grassland, which has been converted to a shrubland by overgrazing. However, at least some shrublands grow on soils developed from limestone or gypsum and so may be partly edaphically controlled.

C. Deserts

There are two main areas of real desert in Africa: the Sahara to the north of the equator and the much smaller Namib of southwestern Africa. Parts of the Horn of Africa are also extremely dry, but only small areas such as the Danakil Depression can be considered as true deserts, if true desert is defined as an area in which plants grow only where there is extra water either from springs or from runoff.

At least some rain falls over much of the Sahara each year. The exception is the eastern end, in the Nile Valley, where no rain may fall for many years in succession. Over much of the rest, enough rain falls in most years to produce a thin ephemeral vegetation in more favorable areas. Grasses are the main component of this vegetation, which dries out quickly to form a standing hay that is a valuable food resource for both wild animals and the domestic animals of nomadic herders.

Formerly several ungulate species were not uncommon in the desert: the addax antelope (*Addax nasomaculatus*) and the scimitar-horned oryx (*Oryx dammah*) being the largest, with several smaller species of gazelle. Most of these have extremely efficient water-conserving strategies and are capable of surviving without drinking. Human population increase, with concomitant competition for space, grazing, and water, and the increased availability of firearms, mean that all of these species are now scarce and endangered. Scattered over the desert are oases where springs from underground aquifers provide year-round water. Most of these oases support permanent human settlements, often dependent on sparse annual crops as well as dates produced by the date palms (*Phoenix dactylifera*) that thrive in these environments.

The Namib Desert in southwestern Africa is also very dry, at least near the coast, but the dryness is mitigated by frequent fogs arising from the cold Benguela Current immediately offshore. Many of the plants and animals in this region survive by collecting water from mist. Furthermore, a few rivers cross the coastal strip, forming linear oases in which water is almost always available. Animals such as oryx (*Oryx gazella*), and even a few herds of elephants, manage to survive in this harsh environment.

One plant family, Mesembryanthaceae, has speciated enormously in this region, and there are hundreds of species. Many of these are very simple in structure, producing clumps of short shoots each bearing just one pair of large thick leaves at any one time. The lower parts of these leaves are buried in the soil, thus avoiding the extreme heat of the surface, and the exposed surface is often translucent, allowing light to reach the buried part of the leaves. The exposed parts of the leaves are also often colored and textured just like the surrounding stones, which camouflages the plants; the members of one genus, *Lithops*, are known as "living stones." One truly remarkable plant grows in this area and nowhere else in the World—*Welwitschia mirabilis*. This is not a flowering plant, and neither is it a true conifer. Each plant consists of a thick woody stem, mostly buried in the ground, bearing two huge strap-shaped leaves that grow continuously from the base, lie on the ground, and wear away at the tips.

D. Montane and Afroalpine Ecosystems

Mountains in Africa are of two kinds: relatively recent volcanics such as Mts. Cameroon, Kilimanjaro, Elgon, and Kenya and the Ethiopian Highlands, and old upthrust portions of basement complex such as the Ruwenzori Mountains and numerous lower ranges in Tanzania such as the Usambara, Uluguru, and Uzungwa Mountains.

Mountain slopes were originally forested, but because of their fertile soils in many places the forests have been cleared for cultivation. However, the higher forests tend to be cool and misty and much less attractive for agriculture, and some of these survive. Trees such as *Podocarpus* and the mountain bamboo *Arundinaria alpina* are common. Epiphytic bryophytes, ferns, and orchids are abundant. Above the montane forest is a belt of "giant heath" formed from large species of *Erica* that attain 10 m in height. As in the montane forest, epiphytic mosses and liverworts are very abundant, often forming dense mats covering the trunks and branches of the giant heaths.

Above the heath zone lies the afroalpine zone proper. The climate here has been described as "summer every day and winter every night." Hot sunny days (this region is often above the cloud zone) are followed by nights during which the temperature plunges to well below freezing, and frost and ice forms on the ground. Here giant groundsels (*Dendrosenecio*) and giant *Lobelia* grow scattered in a low shrubland of everlastings (*Helichrysum*) and *Alchemilla*. The giant groundsels and giant lobelias consist of rosettes of huge leaves borne at the ends of sparsely branched stems. (A very similar growth form is found in the genus *Espeletia* of the Andean paramos.) During the night, the leaves fold upwards to form a dense mass around the delicate terminal buds, thus protecting them from the cold. Water and mucilage accumulate in the rosette and also help to prevent the buds from freezing. The old leaves of some species do not fall, but accumulate as an insulating blanket around the stems; other species have thick corky bark that functions in a similar way.

Animals are generally sparse in montane forests and in the afroalpine zone. Various antelopes move up here from the surrounding savannas, as do elephants and the occasional buffalo. With these come predators, particularly leopards. Specifically montane animals include the mountain nyala (*Tragelaphus buxtoni*) of the heath zone in Ethiopia, and the giant forest hog, a huge, hairy, black pig that occurs in the montane forests of Kenya and elsewhere, but also at lower altitudes in western Uganda. Montane forests are rich in bird species, some of them occurring nowhere else. The giant lobelias are visited and pollinated by sunbirds, which are specialists on them.

E. Wetlands

The wetlands of Africa may be seasonal or permanent. Warping and subsidence of the earth's crust has produced several extensive wetland areas, such as the inland delta of the Niger River in Mali, the Sudd of the Nile River in the Sudan, the inland delta of the Okavango in Botswana, and the Bangweolo swamps of northern Zambia. Many permanent swamps are dominated by the giant sedge *Cyperus papyrus*. Papyrus requires at least some water movement and a reasonable supply of nutrients to thrive; when it does, it attains heights of 3 m or more. Papyrus swamps are very species poor. There are few other plants, most of them climbers like the purple morning-glory, *Ipomoea*, and the yellow-flowered pea, *Vigna luteola*. A swamp antelope, the sitatunga (*Tragelaphus spekei*), shelters in papyrus but finds little food there. Some birds, such as the golden-crowned gonolek (a shrike) and Carruthers' cisticola (a small brown warbler) spend their whole lives in papyrus swamps, but most of the herons and other waterbirds are found along channels and at the swamp margins.

Temporary pools are a common feature of the seasonal region of Africa. A pool will fill early in the rainy season, sometimes from rain and sometimes from overspill from rivers. Fish usually reach all but the most isolated pools; the air-breathing African lungfish can survive in a mucous cocoon in the mud for many months, emerging when water returns. Catfish (*Clarias*) can move over land through wet grass, and species of killifish survive as resistant eggs in the mud, hatching when wetted.

Floating plants like the Nile cabbage (*Pistia*) and water lilies (*Nymphaea*), as well as submerged plants such as bladderworts (*Utricularia*) and hornwort (*Ceratophyllum*), appear from seed in the mud. In game reserves and parks hippopotamuses (*Hippopotamus amphibius*) often move into the pools and churn them into mud, as well as adding nutrients in their droppings. At the end of the wet season the pools start to dry up. Then flocks of herons, egrets, and storks arrive to feed on the fish trapped in the shrinking patches of water, and the hippos move back to permanent water or die.

Seasonally flooded grasslands are found around the edges of many of the swamp regions of Africa. Some of them, like those around the Sudd of southern Sudan, and the Kafue River flats of Zambia, are very large. Among the grasses, species of *Echinochloa* are often common; at least some of these are very tolerant of flooding and their stems can elongate and form a floating mat when water levels are high. Perennial wild rice (*Oryza longistaminata*) is also common. Many of these areas are occupied by pastoral groups such as the Dinka of southern Sudan. They are cattle herders and move seasonally to follow the rise and fall of the flood. There are also wild ungulates; the Nile lechwe (*Kobus megaceros*) is confined to the grasslands at the edges of the Sudd swamps, and the red lechwe (*Kobus leche*) is found in various seasonally flooded grassland areas of Zambia. These animals also follow the rise and fall of the flood.

F. Lakes

Africa has numerous freshwater lakes, some of them very large. They are either shallow or deep lakes. First, there are the deep lakes of the Rift Valleys: Lakes Tanganyika (33,000 km^2, maximum depth of 1460 m), Malawi (704 m), Edward, and Albert. These lakes are permanently stratified. In the temperate zones, low winter temperatures cool the surface waters to 4°C, at which temperature water is densest and sinks, carrying dissolved oxygen to the bottom of the lake. In the tropics there are no significant seasonal variations in temperature, the surface water is always warmer than the depths, and there is little circulation, so that the deeper layers are anaerobic and lifeless. In spite of this, the oxygenated upper layers have a rich fauna and are highly productive. These lakes are ancient, and speciation within them has been rapid. The diversity of molluscs in Lake Tanganyika, for instance, parallels that of the sea, with thick-shelled species occurring on exposed coastal rocks and species with long spines on the shells living on soft mud bottoms. The fish faunas also parallel those of the sea, with some species living mainly near the shore, while others exploit the zooplankton and other food in the open water.

Then there are the shallower lakes, such as Lakes Victoria (69,000 km^2, maximum depth of 90 m), Kyoga, Chad, and Bangweolu. In these lakes the stirring action of the wind is enough to circulate the water and carry oxygen to the lake bottom. Most of the shallower lakes are relatively young in geological terms, but some of the groups of fishes in them have speciated explosively to produce "species flocks," each of whose members exploits a different, often extremely specialized, niche.

In Lake Victoria, which is less than a million years old, the cichild mouth-brooding genus *Haplochromis* is represented by more than 150 species, including plant-eaters, snail-eaters, and fish predators. More bizarrely, there are also species that live by biting off scales and portions of the fins of others, by pulling mayfly larvae from their burrows in dead wood, and by (probably) sucking the eggs and young from the mouths of brooding females. Sadly this astounding diversity is under extreme threat following the introduction of the Nile perch (*Lates niloticus*) to the lake. This is a voracious predator, formerly found only in the Nile system below the barrier of the Murchison Falls, but now introduced to Lake Victoria with the intention of increasing fishing yields. Water hyacinth (*Eichhornia crassipes*) has also been introduced to Lake Victoria and is altering the ecosystem enormously by forming extensive mats at the edge of the lake, which make access for fishing difficult. (These mats may also, however, be providing nursery areas for young fish.) Lake Malawi has a similarly remarkable diversity of cichlid fishes and, so far, neither Nile perch nor water hyacinth has been introduced.

Finally, within the Rift Valley there are a number of lakes that are highly alkaline or saline. Soda (sodium carbonate) is obtained commercially from Lake Magadi in Kenya, and salt (sodium chloride) from Lake Katwe in western Uganda. The soda lakes of the western Rift Valley in Ethiopia, Kenya, and Tanzania, as is often the case with extreme environments, support relatively simple ecosystems in which a few species are present in great abundance. A planktonic blue-green alga, *Spirulina*, often forms almost a soup in the alkaline water. Planktonic copepods (small crustaceans) and chironomid (midge) larvae feed on the algae. Lesser flamingos (*Phoeniconaias minor*) also feed on the algae by filtering water through their beaks. Greater flamingos (*Phoenicopterus ruber*) have more widely spaced filters in their beaks, and feed mainly on the larger copepods and midge larvae, often stirring them out of the mud with their feet. They are not as numerous as the lesser flamingos. Both species migrate up and down the Rift Valley in response to changes in water levels and algal concentration. They breed in huge colonies when water levels are right and food availability is maximal.

G. Coastal Ecosystems

Most of Africa's seashores are sandy or muddy; rock is rare and generally confined to isolated headlands. Coral reefs occur along the east coast from the Red Sea to Mozambique, but not in the west, where cold upwelling water and turbidity virtually exclude them.

The strand lines of sandy shores support open communities of creeping plants such as *Ipomoea pes-caprae* (Convolvulaceae), *Canavalia rosea* (Fabaceae) and *Remirea maritima* (Cyperaceae). The first two have large seeds that float in seawater and are probably thus dispersed; both species occur widely in the Old World tropics. On the landward side of this grow salt-resistant bushes of species such as *Sophora inhambanensis* and *Pemphis maritima*, often forming dense thickets. Such thickets may grade into forest, or, in parts of West Africa, into a wind-swept grassland with bush clumps sculpted by the wind. Some of the grassland plants exist here as prostrate ecotypes that maintain their prostrate growth form in cultivation.

More sheltered shores often support mangrove swamps, particularly in the neighborhood of the mouths of large rivers such as the Niger and the Rovuma (Tanzania). West African mangrove swamps are relatively poor in species; *Avicennia* and species of *Rhizophora* are frequent here. The east coast is richer in mangrove species. The mangrove ecosystem is extremely productive; the breathing roots of the mangroves support communities of seaweeds and oysters, while the mud surface teems with crabs and small fishes. The whole zone is a rich nursery for fish.

Some of the more sheltered sandy flats behind reefs, particularly in northern Kenya and southern Somalia, have extensive beds of seagrasses (*Cymodocea, Halodule*, and *Thalassodendron*). These are not true grasses, although their flat, parallel-sided leaves are similar, but are closer to the pondweeds (Potamogetonaceae). Some grow intertidally, but many extend well below low-tide level and may form extensive lawns on sandy substrates. Here they are grazed by dugongs (*Dugong dugon*). These marine vegetarian mammals are regularly hunted and are becoming scarce and endangered almost throughout their range.

V. MADAGASCAR

Madagascar has been isolated from Africa for at least 140 million years, and from India for around 88 million years. It has a rich flora in which about 80% of the species are endemic, and in at least some cases their closest relatives occur in Southeast Asia rather than Africa. The isolation of Madagascar predated the major adaptive radiation of mammals that has occurred in Africa, and it lacks large grazing mammals—although a dwarf hippopotamus appears to have become extinct less than 1000 years ago—and large carnivores.

At present the largest wild mammal is the bush pig (*Potamochoerus larvatus*), which is believed to be a recent arrival. A group of primitive primates, the lemurs, has radiated into all the habitats on the island. Fossils show that they were formerly even more diverse than they are now; some extinct forms were much larger than any modern species. The largest carnivore is the fossa, a large mongoose-like animal that climbs trees well and is a specialist predator on lemurs. There are also fossil and subfossil remains of giant flightless birds (*Aepyornis*), the last of which seem to have become extinct only a few hundred years ago. Humans reached the island from the east perhaps 1500 years ago. At first their settlements were confined to the coast, but later they spread inland and colonized the central plateau. Humans have had a dramatic effect on the island's natural vegetation and habitats.

Flowering plants were in their earliest stages of evolution when the island became isolated, and a high percentage of species (80% in the legumes, Fabaceae), many genera, and some families are endemic. The vegetation of the island is very diverse. On the eastern side, rainfall is high and there is little or no dry season. Here there are tropical rain forests, very diverse in composition, with no single species being dominant. They differ from the forests of mainland Africa in their lower stature (25–30 m), lack of large emergent trees, the abundance of small palms in the understory, and the frequent occurrence of climbing bamboos. These forests have been considerably reduced by clearance for agriculture, and only scattered fragments remain. Secondary forests are widespread, often characterized by the distinctive traveler's tree (*Ravenala madagascariensis*), with a single stem crowned by huge leaves arranged like a fan in two opposite rows. Higher up the forest takes on a more montane aspect; the trees are shorter and more branched, and epiphytic ferns and mosses are abundant. The highest mountains support a montane thicket of small-leaved ericoid shrubs such as *Erica* (Ericaceae), *Stoebe* (Asteraceae), and everlastings (*Helichrysum*, Asteraceae).

A drier form of forest or woodland seems also to have occupied much of the central plateau, but only tiny fragments remain, and these are under intense pressure from fire, agriculture, and wood cutting for charcoal. The commonest tree is tapia (*Uapaca bojeri*), which may owe its survival to its fire resistance. One legume genus with two species, *Peltiera*, has recently been described from forest fragments in this zone; only three specimens and no living plants are known and it seems likely that the genus was extinct before it was described. The forests of the central plateau have largely been replaced by a species-poor grassland that provides little protection to the soil from erosion so that gullies are widespread and deep.

In the western half of the island, dry deciduous or semideciduous forest survives here and there, particularly in limestone areas, which have often weathered to produce an inhospitable landscape of sharp ridges and pinnacles ("tsingy") that is very difficult of access and unsuited to any kind of agriculture.

The southern end of the island, particularly in the west, is very dry, and here a peculiar thorn forest is found in which the endemic cactus-like family Didieriaceae is common. Lemurs (sifakas) live in this thorn forest. This remarkable vegetation type is threatened by agriculture, particularly sisal cultivation, by grazing, and by cutting for charcoal production. Perhaps because of the absence of large grazing animals, members of several plant families have developed a growth form in which leaves are absent and photosynthesis is carried out in the flattened stems. Several members of the family Fabaceae show this feature.

Another plant growth form perhaps more widely developed in the dry parts of Madagascar than in any other region is the "bottle-tree," in which a thick and swollen trunk supports a rather small crown. The genus *Adansonia*, with one species in Africa (the baobab) and one or two in Australia, has seven species in Madagascar. The flame-tree (*Delonix regia*), now an extremely widespread ornamental tree in the tropics, is one of ten *Delonix* species in Madagascar, with just one other in tropical Africa. Several of the Madagascar species are bottle-trees.

Some of the richest habitats in Madagascar are the rocky outcrops, perhaps because they are sheltered from fires and grazing animals. Numerous endemic species of *Aloe* and succulent spurges (*Euphorbia*), as well as strange single-stemmed spiny succulents (*Pachypodium*), are common on these rocky outcrops and make them striking refuges for the remarkable flora of this isolated island.

VI. THE FUTURE FOR AFRICAN ECOSYSTEMS

What of the future? The human species has been present in Africa for longer than in any other continents. For much of the time, however, humans would have lived as part of the normal ecosystem—essentially part of the wild fauna. Exactly when they began to rise to dominance and to influence the composition and distribution of ecosystems is hard to say. There is some evidence that humans have been using fire (although probably not making it) in Africa for as long as half a million years. Fire is one of the most potent forces for change in tropical vegetation and an increase in fire frequency caused by humans may well have shifted the balance between woody and herbaceous vegetation towards the latter. The southward spread of pastoral peoples, with their knowledge of the power of fire to produce new grass and control trees, seems to have begun by at least 2000 years B.C., perhaps in response to increasing Saharan aridity, and may have been an important force for vegetation change.

Domestication of crops may well have begun at the same time. African plants that have been domesticated include yams (*Dioscorea*), sorghum, upland rice (*Oryza*), and cowpea (*Vigna*), with oil palm (*Elaeis*) in the forest zone. Bananas (*Musa*) probably arrived from Asia in the first millennium A.D., and New World crops such as maize (*Zea*), cassava (*Manihot*), tomato (*Lycopersicum*), and peppers (*Capsicum*) did not arrive until the fifteenth or sixteenth centuries A.D.

Most of the people of Africa survive by subsistence agriculture, or by growing crops that are sold to the rapidly increasing town populations. Shifting cultivation, in which a piece of ground is cleared, cultivated for a few years, and then abandoned for a fallow period of varying length, is the traditional way of exploiting the nutrient-poor soils characteristic of much of Africa. Although this practice is often attacked as wasteful and destructive, it is a very satisfactory mode of land use, so long as population densities remain low and a long fallow period is possible. However, once populations increase, the length of the fallow period falls, as do yields. This increases pressure to find new agricultural land, such as may be opened up in forest by logging activities. The extraction roads that allow people into the forest and the clearings made during timber cutting provide sites for settlement. It is often said that most of the forest loss in Africa is not caused by timber extraction, but rather by the subsequent settlement.

Human populations are increasing throughout Africa, in some nations at alarming rates. It is easy to overlook that an annual increase of 3.5% implies a population doubling every 15 years, and such rates are found in many African countries. Probably a majority of Africans are aware of the problem, but the absence of state care for the old is a considerable incentive to produce numerous children; even if child mortality is high, at least one or two will survive to provide care for the parents in their old age. Improvements in health care tend to come before reductions in birth rates, leading to lower death rates and longer life expectancies.

A rise in population increases the pressure on land, and therefore on natural ecosystems. Cultivators spread farther into areas moist enough for agriculture, and irrigation schemes push out the cultivable boundaries. Pastoralists increase their flocks and herds, which are viewed as cash on the hoof, whose numbers tend to increase until drought or disease cuts them back. Some of the more natural ecosystems, such as those of seasonally dry regions with high wild animal populations, are enclosed in national parks or game reserves, but unless these are very large they are invariably inadequate to accommodate migratory species or those, like the elephant, that normally range over a very wide area. Within the parks and reserves, tourist pressure can also be a problem.

Political instability and the wider availability of firearms are other threats. All of the larger antelopes of the Sahara and its fringes are now endangered because of hunting with guns and motor vehicles, as well as competition with domestic stock for forage and water. The migratory animal populations of southern Sudan have been greatly reduced by hunting with the many firearms now in the region. In the 1970s and 1980s, elephants were hunted throughout Africa for their ivory.

Undisturbed forest is now a rare commodity in the whole of West and East Africa; only in the Congo River Basin are there tracts still in their original state, protected by their size and inaccessibility. Exploitation of forests for timber will no doubt continue and probably intensify, not only for timber for export and internal use but also for charcoal to fuel the cookers of the cities. Additionally, more areas will be cleared for plantation agriculture and for cash crops such as oil palm and cocoa. The woodlands will continue to be heavily exploited for firewood and charcoal, and more favorable sites will be cleared and converted to large-scale agriculture, if soil fertility can be maintained. Fire frequency in the grasslands will continue to increase as the human population increases, although the spread of agriculture tends to reduce fire by fragmenting the area that can be burned.

Yet here and there are signs that change may be taking place. In West Africa there is a tradition of preserving a patch of forest near every village to provide a burial ground, a home for the ancestral spirits, and a source of plants for medicine and poles for building. These sacred groves provide refugia for some forest species. In parts of the Guinea Savanna zone, schemes for sustainable exploitation of the woodlands for firewood and charcoal have been developed and may be taking hold. Overall, however, the future for many plant and animals in Africa, particularly those of the forests, is still bleak.

See Also the Following Articles

DESERT ECOSYSTEMS • FIRES, ECOLOGICAL EFFECTS OF • MEDITERRANEAN-CLIMATE ECOSYSTEMS • NEAR EAST ECOSYSTEMS • TROPICAL ECOSYSTEMS

Bibliography

Cowling, R., and Richardson, D. (1995). *Fynbos. South Africa's Unique Floral Kingdom*. Fernwood Press, Vlaeberg, South Africa.

Cowling, R. M., Richardson, D. M., and Pierce, S. M. (1997). *Vegetation of Southern Africa*. Cambridge University Press, Cambridge, United Kingdom.

Huxley, C. R., Lock, J. M., and Cutler, D. F. (eds.). (1998). *Chorology, Taxonomy and Ecology of the Floras of Africa and Madagascar*. Royal Botanic Gardens, Kew, London.

Kingdon, J. (1990). *Island Africa. The Evolution of Africa's Rare Animals and Plants*. Collin, London.

Kingdon, J. (1997). *The Kingdon Field Guide to African Mammals*. Academic Press, San Diego/London.

Koechlin, J., Guillaumet, J.-L., and Morat, P. (1974, repr. 1997). *Flore et Végétation de Madagascar*. ARG Gautner Verlag, Vaduz, Liechtenstein.

Lawson, G. W. (ed.). (1986). *Plant Ecology in West Africa: Systems and Processes*, John Wiley & Sons, Chichester, United Kingdom.

Lind, E. M., and Morrison, M. E. S. (1974). *East African Vegetation*. Longman, London.

Richmond, M. D. (ed.). (1997). *A Guide to the Seashores of Eastern Africa and the Western Indian Ocean Islands*. SIDA, Stockholm.

White, F. (1983). *The Vegetation of Africa. A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSC Vegetation Map of Africa*. UNESCO, Paris.

AGRICULTURAL INVASIONS

David Pimentel
Cornell University

GLOSSARY

competition Use or defense of a resource, such as a desirable growing site for a plant and a food source or shelter for an animal, by an individual that reduces the availability of the resource for another individual.

competitive exclusion Extirpation or extinction of one species by another in a given area through competition for resources.

exotic (nonindigenous) species A species that occurs in an area outside of its historically known range and that has been introduced into a new habitat or ecosystem, either intentionally or accidentally.

native (indigenous) species A species that occurs naturally in a given area.

THE INTENTIONAL—AND ACCIDENTAL—INTRODUCTION OF VARIOUS BIOLOGICAL SPECIES INTO AGRICULTURE started with the origin of farming about 10,000 years ago. Since then, more than 400,000 species have been moved from one region of Earth to another. Introduced species, like corn, wheat, rice, and other food crops, as well as domestic cattle, poultry, and other livestock, now provide more than 98% of the world's food supply, at a value of more than $5 trillion per year. Other introduced species have been used for landscape restoration, biological pest control, sport, pets, and food processing.

I. INTRODUCTION

Thousands of species have been intentionally introduced to new regions for explicit agricultural purposes, such as the growing of crops and pest control. Yet many species are also accidentally introduced, and these biological "invaders" end up doing harm to the areas they invade. They can cause major economic losses in agriculture, forestry, and other segments of the world economy if crops or other products are destroyed. One recent study reported that the more than 50,000 species that have been introduced into the United States since Columbus' time have resulted in approximately $137 billion in damages (Pimentel *et al.*, 2000a). In addition, introduced species often have substantial negative im-

Encyclopedia of Biodiversity, Volume 1

pacts on global ecological integrity; for example, the presence of an alien invader can result in the extinction of many native species.

It is difficult and complicated to estimate the full extent of the ecological damages to world agriculture caused by exotic species and the number of species extinctions they have caused, because only approximately 20% of the estimated 10 to 15 million species in the world have been described and catalogued. Nonetheless, about 400 of the 958 species on the Threatened or Endangered Species List in the United States are considered at risk primarily because of competition and predation by nonindigenous species. Other species are continually threatened by hybridization with alien species and/or ecosystem changes caused by biological invaders. Estimation of the economic impacts of nonindigenous pests affecting world agriculture is also difficult; however, sufficient data are available to quantify some aspects of the economic impacts on agriculture. This article assesses the magnitude of environmental impacts and some of the economic costs associated with the diverse nonindigenous species that have invaded world agriculture.

II. EXOTIC SPECIES INTRODUCTIONS IN THE UNITED STATES

Most plant and vertebrate animal introductions in U.S. agriculture have been intentional, whereas most invertebrate animal and microbe introductions have been accidental. In the recent past, the rate of and risk associated with exotic species introductions have increased enormously because the human population is growing so rapidly; this ever-increasing number of humans has substantially altered, and continues to alter, the global environment (Pimentel *et al.*, 2000b). In addition, large numbers of people are traveling faster and farther and more goods and materials are being traded among nations all over the world (Bryan, 1996), so the opportunity for even more species introductions is on the rise.

A. Nonindigenous Plants

1. Introduced Crop Plants

Most alien plants now established in the United States were introduced for food, fiber, and/or ornamental purposes. An estimated 5000 plant species have escaped these realms and now exist in U.S. natural ecosystems. This number is substantial when compared with a total of about 18,000 species of native U.S. plants.

Ornamental plants are the largest group of nonindigenous plants that have become established in nature. These plants include those found in cultivated gardens, such as purple loosestrife (*Lythrum salicaria*). Of the approximately 25,000 cultivated plant species, mostly ornamentals brought into Florida, more than 900 have become established in surrounding natural ecosystems (Simberloff *et al.*, 1997). More than 3000 plant species have been introduced into California. New York state also has a large percentage of introduced plant species: 1082 exotics versus 1940 native species.

In crop systems, including forage crops, an estimated 500 introduced plant species have become weed pests. Most of these were accidentally introduced with crop seeds, from ship ballast, or from various imported dead and live plant materials. Some of these accidentally introduced weeds, like yellow rocket (*Barbarea vulgaris*) and Canada thistle (*Cirsium arvense*), have become major weed pests in U.S. gardens and commercial agriculture.

Most of the 5000 nonindigenous plants established in U.S. natural ecosystems have displaced some native plant species. Alien grasses introduced as forage species have diminished the health of natural ecosystem functions as a result of nutrient losses, changes in microclimates, altered vegetation succession, and increased incidence of fires. Nonindigenous weeds are spreading and invading approximately 700,000 ha of U.S. wildlife habitat each year (Babbitt, 1998).

One of these weed pests is the European purple loosestrife, which was introduced in the early nineteenth century as an ornamental garden plant. It has been spreading at a rate of 115,000 ha per year and is changing the basic structure of most of the wetlands that it has invaded. The dominant, monotypic stands of purple loosestrife have reduced the biomass of 44 native plants and endangered the wildlife that depends on these native plants.

2. Economic Costs of Introduced Plants

In U.S. agriculture, weeds cause a reduction of 12% in crop yields; in economic terms, that represents an approximate $33 billion loss in crop production annually, based on the potential crop value of all U.S. crops of more than $267 billion per year. Using the estimate that about 73% of the weeds are nonindigenous, it follows that around $24 billion of these crop losses are due to introduced weeds. In addition, approximately $4 billion in herbicides are applied to U.S. crops, of which about $3 billion can be attributed to nonindigenous weed control.

U.S. pastures provide about $10 billion in forage crops annually, with total losses due to weeds estimated at nearly $2 billion. Since about 45% of these weeds species are nonindigenous, I estimate that forage losses due to these weeds are close to $1 billion each year.

Some introduced weeds are actually toxic to cattle and wild ungulates. For example, leafy spurge (*Euphorbia esula*) contains a toxic substance that, when eaten by cattle, acts as an irritant, emetic, and purgative and can cause dysentery and sometimes death. Some animals, after unpleasant experiences, learn to avoid eating the plant. If the vegetation on grazing land is composed of just 10 to 20% leafy spurge, the cattle will not graze there and the value of the land as forage is lost. Total losses resulting from the prevalence of this weed on western rangelands account for nearly $110 million per year due to toxic effects and reduced grazing by the cattle (this cost is included in the $1 billion in losses mentioned earlier) (OTA, 1993).

Several nonindigenous thistles also replace desirable native plant species in pastures, rangelands, and forests and similarly reduce cattle grazing. According to United States Secretary of the Interior Bruce Babbitt (1998), ranchers spend about $5 billion each year to control invasive nonindigenous weeds in pastures and rangelands; despite those expensive efforts, though, these weeds continue to spread.

Lawn, garden, and golf course management costs about $36 billion per year. A significant proportion of these management costs is related to pest control. In addition to the high cost of this commercial pest control, it is estimated that about $1.3 billion (included in the $36 billion) is spent on outdoor residential weed, insect, and disease pest control each year. Given the high percentages of nonindigenous species, I estimate that about $500 million is spent on residential nonindigenous weed control, and an additional $1 billion is invested in nonindigenous weed control on golf courses. Therefore, the total economic cost of introduced weeds on U.S. cropland, pastureland, and other uses is about $35 billion annually (Table I).

B. Nonindigenous Vertebrates

1. Mammalian Introductions

Introduced mammal species include dogs, cats, horses, burros, cattle, sheep, pigs, goats, monkeys, and deer (Layne, 1997). Several of these animals escaped or were released into the wild, and many have become pests by either preying on native animals, grazing on vegetation, or intensifying soil erosion.

For example, goats (*Capra hircus*) introduced on San Clemente Island, California, are responsible for the extinction of eight endemic plant species and the endangerment of eight other native plant species on the island (Kurdila, 1995). Horses (*Equus caballus*) and burros (*Equus asinus*) released in the western United States have reached populations of up to 50,000 animals. These animals graze heavily on native vegetation, allowing nonindigenous annuals to displace native perennials. Furthermore, burros inhabiting the northwestern United States compete for the primary food sources of native bighorn sheep and seed-eating birds, thereby reducing the abundance of these native animals (Kurdila, 1995).

Overall, large populations of feral horses and burros are costly in terms of lost forage for livestock and wild-

TABLE I

Economic Losses to Biological Pest Invaders in Agriculture in the United States, United Kingdom, Australia, South Africa, India, and Brazil (in Billions of Dollars)

Pest invader	United States[a]	United Kingdom	Australia	South Africa	India	Brazil	Total
Plant weeds	$35.000	$1.200	$2.400	$2.500	$23.724	$12.300	$77.124
Mammals							
Rats	19.000	4.400	1.200	2.700	25.000	3.200	$55.500
Other	1.310	1.200	4.380	—	—	—	$6.890
Birds	2.100	0.270	—	—	—	—	$2.370
Arthropods	15.000	0.840	1.000	1.700	10.200	4.600	$33.340
Microbes							
Plant pathogens	23.500	1.700	2.700	2.900	21.400	18.300	$70.500
Livestock	—	—	0.021	1.000	—	—	$1.021
Total	$95.910	$9.610	$11.701	$10.800	$80.324	$38.400	$246.745

[a] Includes pest control costs for lawns, gardens, and golf courses. See text for details.

life. Based on a minimum of $10 per animal-unit-month (AUM) per hectare, and assuming 10 ha of pasture per animal per year, these large animals exact damage of at least $5 million per year. This cost does not include the extensive environmental damage to ecosystems, which can be up to 10 times this amount when increased soil erosion and the reduction of native plants and associated native animals are considered.

Pigs (*Sus scrofa*), which are native to Eurasia and North Africa, have been introduced into some U.S. parks (like the California coastal prairie) for hunting, and have substantially changed the vegetation in those places. Feral pigs are also a serious problem for U.S. agriculture. In Hawaii, more than 80% of the soil is bare in regions inhabited by pigs (Kurdila, 1995). This disturbance allows annual plants to invade the overturned soil and intensifies soil erosion. Pig control per park in Hawaii, with about 1500 pigs per park, costs about $150,000 each year (based on $100 spent per pig for control). Assuming that the three parks in Hawaii have similar pig control problems, the total is $450,000 per year just in Hawaii (R. Zuniga, personal communication).

Pigs also have become a serious problem in Florida, where the feral pig population has risen to more than 500,000. In a few cases these invaders are beneficial—for example, pigs serve as a food source for the endangered Florida panther (Layne, 1997). For the most part, though, in Florida and elsewhere, pigs damage hay, grain, peanut, soybean, cotton, and various vegetable crops. Pigs also damage fences, livestock watering devices, ponds, and young livestock. In addition, they transmit and are reservoirs for serious diseases that affect both humans and livestock, like brucellosis, pseudobrucellosis, and trichinosis.

Nationwide, there are an estimated 4 million feral pigs (Pimentel *et al.*, 2000a). Control costs for feral pigs range from $36 per pig for poison control to $140 per pig for hunting with dogs. Estimated damages are at least $300 per pig per year. Assuming that 4 million feral pigs inhabit the United States, the yearly cost is about $800 million per year. This is a conservative estimate because pigs cause significant environmental damage that is not easily translated into dollar values.

In addition to large animals like horses and pigs, many small mammals have been introduced both accidentally and intentionally into the United States. Some examples of these species are the European (black or tree) rat (*Rattus rattus*), Asiatic (Norway or brown) rat (*Rattus norvegicus*), house mouse (*Mus musculus*), nutria (*Myocaster coypus*), European rabbit (*Oryctolagus cuniculus*), cat (*Felis cattus*), and dog (*Canis familiaris*) (Layne, 1997).

Some rodents have become serious pests on farms, in industries, and in homes (Layne, 1997). On farms, rats and mice are particularly abundant and destructive. On poultry farms there is approximately 1 rat for every 5 chickens (D. Pimentel, unpublished data). Assuming this ratio, the total rat population on U.S. poultry farms may easily number more than 1.4 billion. Even with a decline in the rat population since these observations were made, I estimate that the number of rats on poultry and other farms is still approximately 1 billion. Given an estimated 1 rat per person, there are an estimated 250 million rats in U.S. urban and suburban areas. All told, there is a total of around 1.25 billion rats in the United States.

Various studies report that an individual rat consumes and/or destroys from 210 to 740 kg of grain each year. If it is conservatively assumed that each rat consumes and/or destroys grain and other materials valued at $15/yr, then the total cost of destruction by rats in the United States is more than $19 billion per year. In addition, rats cause fires by gnawing electric wires, pollute foodstuffs, and act as major vectors in the transmission of several diseases, including salmonellosis and leptospirosis, and to a lesser degree plague and murine typhus.

The Indian mongoose (*Herpestes auropunctatus*) was first introduced into Jamaica in 1872 for the biological control of rats in sugarcane plantations. It was soon introduced to Puerto Rico, other West Indian Islands, and Hawaii. The mongoose was effective in reducing the Asiatic rat population in some areas, but with reduced competition the European rat returned to being a major pest in sugarcane fields. Moreover, the mongoose preyed heavily on ground-nesting birds and subsequently reduced their numbers. It also preyed on amphibians and reptiles that were beneficial to biological pest control, resulting in a minimum of 7 to 12 extinctions in Puerto Rico and other islands of the West Indies. In addition, the mongoose emerged as the major vector and reservoir of rabies and leptospirosis in Puerto Rico and other islands. Based on public health effects, poultry losses, extinctions of amphibians and reptiles, and the elimination of many native birds, we estimate that the mongoose causes approximately $50 million in damages each year in Puerto Rico and the Hawaiian Islands (D. Pimentel, unpublished data; R. Zuniga, personal communication).

Finally, most dogs introduced into the United States have been domesticated, although some have escaped into the wild. Some of these feral dogs run in packs and kill deer, rabbits, and domestic cattle, sheep, and goats. It is estimated that feral dog packs in Texas cause more than $5 million in livestock losses each year. Dog packs have also become a serious problem in Florida (Layne, 1997). Assuming $5 million in damages for the other 47 continental states, total losses in livestock kills per year would be approximately $10 million per year.

2. Exotic Birds

Of the 97 bird species introduced into the United States, only 5% are considered beneficial; a majority (56%) are pests (Temple, 1992). Several nonindigenous bird species, including chickens and pigeons, were introduced into the United States for agricultural purposes.

The common myna (*Acridotheres tristis*), introduced into Hawaii for biological control, helped control pest cutworms and armyworms in sugarcane. However, it became the major disperser of seeds of the introduced pest-weed *Lantana camara*. To cope with the weed problem, Hawaii was forced to introduce insect biocontrol agents for the weed (Kurdila, 1995).

The English or house sparrow (*Passer domesticus*) was introduced into the United States intentionally in 1853 to control the canker worm. By 1900, the birds themselves were considered pests because they damage plants around homes and public buildings and consume wheat, green corn, and the buds of fruit trees. About 70% of the house sparrow's diet consists of grains (the introduced alfalfa weevil accounts for the other 30%). Furthermore, English sparrows harass robins, Baltimore orioles, yellow-billed cuckoos, and black-billed cuckoos. They also displace bluebirds, wrens, purple martins, and cliff swallows from their nest sites, and are associated with the spread of about 29 diseases that affect both humans and domestic livestock.

Similarly, European starlings (*Sturnus vulgaris*) are serious pests and are estimated to occur at densities of more than 1 per hectare in agricultural regions. They are capable of destroying approximately $2000 worth of cherries per hectare. In grain fields, starlings consume about $6 worth of grain per hectare. Therefore, assuming approximately $5/ha for all damages to agriculture crop production in the United States, the total economic loss due to starlings would total approximately $800 million per year. In addition to these economic damages, starlings also have some negative environmental impacts. For example, starlings are aggressive and have displaced numerous native bird species, which can disrupt precariously balanced ecosystems. Starlings also have been implicated in the transmission of 25 diseases, including parrot fever and other diseases in humans and livestock.

The domestic pigeon (*Columba livia*) has been introduced for agricultural production and it has invaded most cities of the world, including U.S. cities. Pigeons are considered a nuisance because they foul buildings, statues, cars, and sometimes pedestrians with their droppings, and because they feed on agricultural grain. Each pigeon consumes an average of 15 kg of grain per year (Smith, 1992). The economic impacts caused by fouling are estimated to be at least $9 per pigeon per year (based on $9 in control costs per pigeon). Assuming there is 1 pigeon per hectare in urban areas or approximately 0.5 pigeon per person in urban areas, common pigeons cause at least $1.1 billion in damages each year. This estimate does not take into consideration pigeons' role as reservoirs for over 50 serious human and livestock diseases, including parrot fever, ornithosis, histoplasmosis, and encephalitis.

Thus, if we assume $800 million per year in eco omic losses from starlings, $1.1 billion per year from pigeons, and $200 million for house sparrows and other birds, the damages from nonindigenous pest birds are estimated to be $2.1 billion per year.

3. Nonindigenous Amphibians and Reptiles

The cane toad (*Bufo marinus*) and bull frog (*Rana catesbeiana*) were introduced into Florida, Puerto Rico, Hawaii, and other warm regions, in some cases for the biological control of pest insects. However, the cane toad, with its toxic skin glands, has proved lethal to birds, dogs, cats, and other mammals.

C. Nonindigenous Invertebrates and Microbes

Approximately 4600 arthropod species (2600 species in Hawaii and more than 2000 in the continental United States) have been introduced into the United States. Eleven earthworm species and nearly 100 aquatic invertebrate species also have been introduced (OTA, 1993). More than 95% of these introductions were accidental—many invertebrate species gained entrance via plant introductions, in soil, or in water ballast from ships.

1. Insects and Mites

Approximately 1000 nonindigenous insect and mite species are pests in crops, stored-food products, and structures. Hawaii has about 5200 identified native insect species, and an additional 2600 introduced insect species (Howarth, 1990). Introduced insects account for 98% of the pest insects in that state. In addition to

Florida's 11,500 indigenous insect species, 949 immigrant species have invaded the state (42 species were introduced for biological control). In California, the 600 introduced species are responsible for 67% of all crop losses. Some of the California pests include the cottony cushion scale (*Icerya purchasi*) and alfalfa weevil (*Hypera postica*).

Each year pest insects destroy about 13% of potential crop production at a cost of about $33 billion in U.S. crops. Considering that about 40% of the pests were introduced, we estimate that these alien pests cause about $13 billion in crop losses each year. In addition, about $1.2 billion in pesticides are applied for all insect control each year. The portion applied against introduced pest insects costs approximately $0.5 billion. Therefore, the total cost for introduced nonindigenous insect pests is approximately $13.5 billion annually.

Lawn, garden, and golf course management activities cost about $36 billion annually. More than $3 billion per year is related to the expenses of pest control. Assuming the presence of 40% nonindigenous insect pests, I estimate the control costs in lawns, gardens, and golf courses to be about $1.5 billion each year.

Other introduced insect species have become pests of livestock and wildlife. For example, the imported red fire ant (*Solenopsis invicta*) kills poultry chicks, lizards, snakes, and ground-nesting birds. In some areas fire ants are extremely abundant, with as many as 367 nests per ha. The estimated damage to livestock, wildlife, and public health caused by fire ants in Texas is estimated to be $300 million. An additional $200 million per year is invested in controlling these ants. Assuming equal damages in other such infested southern states, the fire ant damages total approximately $1 billion per year.

2. Nonindigenous Earthworms

There are approximately 70 native U.S. earthworm species. In disturbed habitats, 11 species of introduced earthworms have reduced the numbers of some of the native species. Although detailed data are not available concerning impacts of nonindigenous earthworms on U.S. earthworm species and soil quality, earthworms are generally beneficial to soil productivity and formation in agriculture.

3. Microbe Introductions in Agriculture

Although some microbes were intentionally introduced into the United States for wine and cheese making, most were accidentally introduced and some have become serious pests. The number of microbe species introduced into the United States cannot be estimated because several thousand species can exist in a single gram of soil. Only a small fraction of all of these species have been identified.

More than 100 species of microbes have been intentionally introduced for processing wine, beer, cheese, and other foods. In addition, about 50 microbes have been introduced for the biological control of pest insects. A strain of *Bacillus thuringiensis* (BT) was introduced and has been used extensively to control pest caterpillars. Other strains of BT have been developed to control beetles and mosquitoes.

An estimated 121 species of microbes, mostly introduced inadvertently in seeds and other parts of host plants, have become major crop pests in the United States. U.S. crop losses to all plant pathogens total approximately 12% of crops planted, a loss of $33 billion per year. Approximately 65% of these crop losses—an estimated $21 billion per year—are attributable to nonindigenous plant pathogens. In addition, $0.72 billion is spent annually on fungicides to control plant pathogens; approximately $0.5 billion of this goes toward the control of nonindigenous plant pathogens.

As mentioned earlier, lawn, garden, and golf course management activities have an annual cost of about $36 billion; most of this is spent for pest control. Assuming that about $3 billion of management costs are related to plant pathogen control and 65% of the pests are nonindigenous pathogens, we estimate the costs caused by introduced plant pathogens in lawns, gardens, and golf courses to be $2 billion each year. Therefore, the damages caused by nonindigenous pathogens and the attempts to control them total about $23.5 billion annually.

When all of the foregoing economic costs in crop and pastureland losses and for pest control for nonindigenous weeds, mammals, birds, insects, and plant pathogens are combined, the annual total is approximately $96 billion to the U.S. economy.

III. EXOTIC SPECIES INTRODUCTIONS IN THE UNITED KINGDOM

A. Nonindigenous Plants

There are a total of 26,000 introduced plant species in the United Kingdom. An estimated 25,000 species are established in U.K. botanical gardens alone, 14,000 species are cultured as commercial horticultural crops, and there are 6000 species of noncultivated aliens. There are 1169 species of exotic plants known to be established in natural ecosystems. The native flora of the United

Kingdom numbers only 1515 species (Crawley *et al.*, 1996).

Most of the alien plants occur in relatively few types of habitats. More than 80% of the alien species are present in waste ground areas, urban sites, roadway sides, and agricultural habitats. An estimated 63% of the alien plants occur in hedges and scrub areas; croplands and gardens harbor about 43% of the alien species (Crawley *et al.*, 1996). Finally, about 40% of alien species occur in rock walls and woodlands. It is interesting that plant communities like grazed, mesic grasslands and native *Pinus sylvestris* woodlands contain no alien plant species.

In U.K. agriculture, weeds cause an average reduction of about 10% in crop yields; however, this loss can be as high as 32% in some crops. In economic terms, about $2.8 billion in potential crop production is annually lost due to weed infestations. Given the estimate that about 43% of the weeds are alien, it follows that $1.2 million of the crop losses are due to introduced weeds (see Table I).

B. Nonindigenous Vertebrates

1. Introduced Mammals

The total number of mammalian species in the United Kingdom is 54; 17 of these are alien species. The mammal introductions include domestic animals like dogs, cats, cattle, horses, sheep, and pigs, as well as other nondomesticated mammals. Some of the species that were intentionally or accidentally introduced into the United Kingdom include the gray squirrel (*Sciurus carolinensis*), European rabbit, North American mink (*Mustela vison*), brown rat, and black rat (Lever, 1994). All of these animals, except for cattle and horses, have escaped into the wild and are now well established.

In the United Kingdom, rodents are serious pests on farms, in industries, and in homes. The estimated number of rats are estimated at 5 per person (based on the rat-to-person ratio in the United States), including rats on farms (Pimentel *et al.*, 2000a). Thus the total number of rats is 295 million. Using the same value of $15 for food and other goods damaged per rat, the cost of rats in the United Kingdom is about $4.4 billion annually.

The European rabbit is also abundant in the United Kingdom, with densities in some areas of up to 30 rabbits per hectare. Assuming approximately 10 rabbits/ha on the 7 million ha of cropland in the United Kingdom with an estimated $11 damage per rabbit, the total annual economic damages from European rabbits are $800 million/yr. They are reported to reduce wheat production by about 5 to 8%, and to reduce forage production for livestock by about 20%. With 11 million ha of pasture for livestock, we estimate these losses to be $400 million/yr. Thus, the total damage from the European rabbit in the United Kingdom is $1.2 billion/yr.

2. Introduced Birds

Out of the 542 bird species in the United Kingdom, 47 are alien species. Some of the introduced species include the Canada goose (*Branta canadensis*) and the little owl (*Athene noctua*). Only one of the 47 introduced bird species is currently causing major ecological and/or economic problems—the common pigeon. This bird does invade agriculture, but is most common in cities and towns (Lever, 1994). Pigeons are a particularly serious problem for reasons discussed earlier, namely, the pollution of city surfaces with their droppings and their consumption of grain in agriculture.

Assuming there is 1 pigeon per hectare, or 0.5 pigeon per person in urban areas (as in the United States), then there are approximately 30 million pigeons in the United Kingdom. The estimated damage that a pigeon causes is a minimum of $9 per year to crops and structures in cities and towns (Pimentel *et al.*, 2000a). Therefore, pigeon damages are estimated to be at least $270 million/yr. This does not include the role of pigeons as reservoirs for over 50 human and livestock diseases, including parrot fever, ornithosis, histoplasmosis, and encephalitis. Pigeons are also responsible for transmission of at least three diseases to U.K. poultry, including Newcastle disease.

C. Nonindigenous Invertebrates and Microbes

1. Introduced Arthropods

There are approximately 23,000 native species and 1700 introduced species of arthropods in the United Kingdom. An estimated 1500 species are of economic importance; about 169 alien species are considered pests. An estimated 30% of the crop losses in the United Kingdom are associated with the introduced arthropod pests, as compared with about 40% in the United States (Pimentel *et al.*, 2000a).

Arthropods damage and/or destroy approximately $2.8 billion in crops in the United Kingdom each year, based on the average of 10% crop losses per year. With about 30% of these losses due to introduced arthropods, they cause an economic loss of about $840 million/yr.

2. Introduced Crop Plant Pathogens

An estimated 74% of the plant pathogens in the United

Kingdom were introduced when seeds and other crop parts were brought into the country for agriculture. Approximately $8 billion is lost to all pests in crop production; about 8% of total potential production is lost to plant pathogens at a cost of about $2.3 billion/yr. If 74% of these crop losses are due to introduced plant pathogens, then about $1.7 billion/yr is associated with introduced microbes in crops.

IV. EXOTIC SPECIES INTRODUCTIONS IN AUSTRALIA

A. Nonindigenous Plants

There are approximately 20,000 vascular plant species in Australia (D. Pimentel, unpublished data), including an estimated 1952 alien species. The rate of alien species introductions into Victoria, Australia, alone has been 5 to 6 species per year during the past century. Many of these species have become weeds and have invaded a wide range of environments. The invasive plants are a serious problem in both agricultural and wild ecosystems, where they disrupt key natural ecosystems, alter fire regimes, and reduce the resources for native animals. An estimated 60% of the weed species in crops in Australia are alien (based on a survey of major weeds in cereal crops).

The introduced blackberry (*Rubus proceus*) alone causes $77 million worth of damages to crop production each year. The indirect and direct losses due to all weeds in pastures are estimated to be $970 million/yr. Weeds are estimated to cause about $4 billion in total damages in crops and pastures. Therefore, invasive plants cause approximately $2.4 billion/yr in losses to agriculture (see Table I).

B. Nonindigenous Vertebrates

1. Introduced Mammals

There are presently 20 introduced mammals in Australia, including the European rabbit, hares (*Lepes europaeus*), water buffaloes (*Bubalis bubalis arnee*), cats, dogs, foxes (*Vulpes vulpes*), sheep, goats, cattle, horses, camels (*Camelus dromedarius*), pigs, and donkeys. In comparison, the number of native mammals is 227. The populations of some of these introduced mammals in the Northern Territory are quite high: buffaloes, 340,000; horses, up to 300,000; donkeys, up to 140,000; and camels, up to 30,000. The populations of these herbivores are much too high, but the costs of implementing a control program are also very high. For example, the cost of controlling buffalo is nearly $30 per animal.

The herbivorous animals significantly reduce the vegetative cover by overgrazing and this intensifies soil erosion and encourages annual plants and inedible shrubs in pastures (Lever, 1994). Surprisingly, the most serious pest herbivorous mammal in Australia is not one of these large-hoofed herbivores, but rather the European rabbit. In Tasmania, the rabbit population in 1952 reached a density of up to 250 rabbits/ha. The total number of rabbits in Australia ranges from 200 to 300 million. Approximately 15 rabbits consume the equivalent pasture forage of one sheep (Emmerson and McCulloch, 1994). The impact of rabbits on sheep production per year is estimated to be $110 million, including reduced sheep production and rabbit control costs. If we assume a conservative estimate of only 0.5 rabbit/ha for cropland and pastureland and also assume that each rabbit causes a minimum of $5 damage, then on the 465 million ha of Australian cropland and pastureland, rabbits are causing at least $1.2 billion/yr in damages.

Feral pigs are also a serious problem in Australia, as they damage fences and spread animal diseases, including tuberculosis, brucellosis, rabies, and foot-and-mouth disease (Lever, 1994). The number of feral pigs ranges from 4 to 20 million. These pigs damage crops, kill lambs, and damage the natural environment; they are estimated to cause at least $80 million/yr in damages (Emmerson and McCulloch, 1994).

The house mouse damages crops, houses, farm machinery, and livestock production. Estimates for these annual losses range from $50 to $100 million/yr. Far more serious than the house mouse, though, are the invading brown and black rats. In the United States, there are 4.6 rats per person (Pimentel *et al.*, 2000a). Assuming the same number of rats per person in Australia and that each rat causes $15 of damage, then for a human population of 18 million, rats are causing approximately $1.2 billion/yr in damages in Australia.

Introduced pet cats and feral cats are also a serious problem, especially for native bird, mammal, and amphibian populations. There are an estimated 18 million feral cats in Australia and each cat is estimated to kill 8 birds per year. Assuming that 144 million birds per year are killed, and that each bird has a value of $30, then damages from cats alone reach about $4.3 billion/year.

2. Introduced Birds

Australia has 850 bird species, of which about 70 are alien species. Most of the introductions have been intentional, including the English starling, English house sparrow, and common pigeon. These birds are often restricted to cities and towns. Pigeons cause similar problems in Australia as they do elsewhere.

3. Introduced Reptiles and Amphibians

There are about 700 species of reptiles and amphibians in Australia, but only 2 of them are nonindigenous. One of these introduced amphibians, the cane toad (*Bufo marinus*), was introduced as a biological control agent for insect pests in sugarcane fields. Unfortunately, the toad has become a pest itself because it is poisonous to dogs, cats, and other mammals that attack it.

C. Nonindigenous Invertebrates

1. Introduced Arthropods

There are an estimated 108,000 arthropod species in Australia, with 54,000 species of native and nonnative insects and mites and 10,000 mollusk species. Crop losses due to insects and mites are estimated at 10.7% of Australia's gross potential production of $22 billion/yr. An estimated 36% of the pest arthropods in Australia are alien species. Based on arthropod-caused crop losses of about $2.4 billion/yr, the exotic pests account for losses of $860 million/yr. In addition, three exotic insects and mites cause $228 million/yr in damages to the wool industry alone. Thus, we estimate that exotic insect and mite species in Australia cause losses of at least $1 billion/yr.

2. Introduced Crop Plant Pathogens

If the total potential crop production in Australia is $22 billion/yr and about 15.2% of crop losses are due to plant pathogens, the economic costs of these pathogens total about $3.3 billion/yr. Because a large number of plant pathogens are introduced with crop seeds and other plant parts, an estimated 82% of the plant pathogens in crops are believed to be alien species (based on plant pathogens in field crops). Therefore, Australia loses $2.7 billion/yr in crops from exotic plant pathogens.

3. Microbe Introductions Affecting Livestock

There are an estimated 44 exotic animal diseases that could infect livestock in Australia. One exotic livestock disease, sheep pox, costs the wool industry an estimated $21 million/yr (D. Pimentel, unpublished data).

V. EXOTIC SPECIES INTRODUCTIONS IN SOUTH AFRICA

A. Nonindigenous Plants

There are 24,000 plant species in South Africa, including an estimated 8750 alien species. Most of the alien species were introduced from South America and Australia and have in turn invaded a wide range of environments. A total of 273 species of introduced plants in South Africa are serious weeds in crops (Bromilow, 1995).

Reduced crop production due to all weeds is 16.6% of potential crop production and totals $3.7 billion/yr in losses (total potential agricultural production is $22 billion/yr). Assuming that 67% of the weeds in crops are alien (Bromilow, 1995), then the total loss in crop production due to alien weeds is $2.5 billion/yr (see Table I).

Two of the most serious plants invading pasturelands include the shrub *Lantana camara* and the cactus plant *Opuntia ficus-indica*. In addition to these two weeds, there are approximately 800 alien weed species out of a total of 1604.

B. Nonindigenous Vertebrates

1. Introduced Mammals

There are 16 species of introduced mammals in South Africa, including the European rabbit, hares, water buffaloes, cats, dogs, sheep, goats, cattle, horses, camels, pigs, and donkeys. The total number of mammal species in South Africa, including alien species, is 247. As in Australia, the herbivore populations in South Africa are much too high for the resources that are available, and they significantly reduce the vegetative cover by overgrazing. This intensifies soil erosion and encourages some annual plants, weeds, and inedible shrubs to take over the pasturelands (Pimentel *et al.*, 2000b).

Feral pigs are also a serious problem; they damage fences and spread disease in South Africa much as they do in Australia. The estimated control cost per feral pig in the United States is about $100 (Pimentel *et al.*, 2000a), about the same as in South Africa.

Rats are a serious problem in South African agriculture as well as in urban areas. There are an estimated 4.6 rats per person in South Africa (based on data from the United States; Pimentel *et al.*, 2000a), which has a human population of 39 million people. Thus, there are about 179 million rats in South Africa; each rat is assumed to cause $15 in damages per year. From these estimates, losses from rat damage total $2.7 billion/yr.

2. Introduced Birds

Of the 725 bird species in South Africa, only 8 are alien species. Some of the introduced species include the English starling, common pigeon, Indian myna, and English house sparrow. Only the starling and pigeon are currently causing economic problems, and their damage is restricted to cities and towns. The problems with starlings and sparrows are very similar to the negative effects that these birds have elsewhere; namely, they pollute structures with their droppings and cause agricultural losses by consuming grains and fruits. In addition, pigeons, starlings, and sparrows are known reservoirs and vectors of up to 50 different diseases of humans and livestock.

C. Nonindigenous Invertebrates and Microbes

1. Introduced Arthropods

An estimated 80,000 species of insects, 6000 species of spiders, and numerous other arthropod species exist in South Africa. About 20% of these are believed to be exotic. One of the most serious invaders is the Argentine ant (*Linepithema humile*), which causes major problems because it destroys native vegetation, including endangered plants. The same ant also negatively affects native ants and other species of arthropods by competing with them for limited resources.

Insect and mite pests in agriculture cause 16.7% or \$3.7 billion in losses of potential crop production each year. Because approximately 45% of the insect and mite pests are exotic, the economic losses to exotic pests are estimated to be \$1.7 billion/yr.

2. Introduced Crop Plant Pathogens

Approximately 85% of the plant pathogens that attack crops in South Africa are nonnative species (based on an assessment of diseases of fruits and vegetables). Most of these pathogens were introduced with the introduction of crops into South Africa. In forests, around 69% of the pathogens are exotic. Plant pathogens in South Africa cause an estimated 15.6% or \$3.4 billion/yr loss of the potential crop production. Since 85% of the pathogens are exotic, crop losses to exotic species total \$2.9 billion/yr.

3. Microbe Introductions Affecting Livestock

Several serious livestock diseases, including tuberculosis, brucellosis, East Coast fever, anthrax, and rinderpest, infect livestock and other animals in South Africa (D. Pimentel, unpublished data). Estimates suggest that these livestock diseases are causing losses of around \$1 billion/yr (Table I).

VI. EXOTIC SPECIES INTRODUCTIONS IN INDIA

A. Nonindigenous Plants

There are 45,000 plant species in India, and an estimated 18,000 of these are alien species (Saxena, 1991). Many of the alien species have become weeds and have invaded a wide range of environments. Several weed species have been introduced along with the introductions of new crops. Weeds are estimated to cause a 30% loss in potential crop production, which totals about \$54 billion/yr. Assuming that 42% of the weeds in crop production are alien, the total cost associated with invading weeds is \$22.8 billion/yr (see Table I).

Lantana camara, a shrub introduced from South America as an ornamental plant, is a major weed in India, and it has invaded most pasturelands (13.2 million ha). *Lantana* is toxic to cattle and the cost of controlling it is \$70/ha. Only about 4% of India's land area is in pasture, yet total damage per year from *Lantana* is \$924 million.

B. Nonindigenous Vertebrates

1. Introduced Mammals

There are approximately 30 species of nonnative mammals in India, including cats, dogs, sheep, goats, pigs, axis deer (*Axis axis*), house mice, and rats. The total number of mammal species in India, including alien species, totals 320 (Pimentel *et al.*, 2000b). As in other countries, the introduced herbivorous animals significantly reduce the vegetative cover by overgrazing, resulting in intensified soil erosion and increased invasion of pastureland by inedible weeds and shrubs.

Rats number at least 2.5 billion in India, or about 2.7 rats per person (Pimentel *et al.*, 1999). They attack crops in the field and are estimated to reduce potential crop yields by about 2%. In addition, rats are especially serious pests of stored grain supplies. Various studies report that an individual rat consumes and/or destroys about 210 kg of grain per year in India, and up to 740 kg of grain per year in Pakistan. Rats are estimated to destroy about 12 million tons of grain per year in India (D. Pimentel, unpublished data), in addition to damaging other foods, goods, and structures. In India I estimate that each rat causes at least $10 of damage per year, and thus they are responsible for at least $25 billion annually (Pimentel *et al.*, 2000b). In addition, rats are major vectors of and carriers of 38 human and livestock diseases. An average of 250,000 people die each year from the plague in India.

2. Introduced Birds

The number of bird species in India is 1221, including migrants and vagrants. Introduced species number only 4, and include the English house sparrow, common pigeon, black francolin, and Alexandrine parakeet. In neighboring Pakistan, sparrows reduce potential wheat yields by 170,000 tons at a cost of $26 million/yr (D. Pimentel, unpublished data). Based on these data, losses in India are calculated to be $50 million/yr. Also, the common pigeon is a problem in agriculture, as well as in cities and towns, where it consumes grains and fouls buildings, statues, cars, and the occasional unlucky pedestrian. Pigeons are also involved in the spread of about 50 diseases that affect humans and livestock.

C. Nonindigenous Invertebrates and Microbes

1. Introduced Mollusks

An estimated 1500 species of land mollusks exist in India; several of these are exotic. The giant African snail (*Achatina fulica*) destroys from 0.29 to 4.3 g of grain per day per snail from eight major crops.

2. Introduced Arthropods

There are more than 30,000 species of arthropods in India. An estimated 560 mite species and about 600 species of insects are crop pests. About 30% of the pest species are introduced arthropods, and as a group they reduce potential crop production by 18.7%. Based on total potential crop production in India of $181 billion/yr, crop losses to alien arthropods total $10.2 billion/yr.

3. Introduced Crop Plant Pathogens

Plant pathogens reduce potential crop production in India by approximately 16%, at a total cost of $29 billion/yr. There are approximately 30,000 species of plant pathogens in India, and about 74% of the major pathogens are exotic species (based on the major plant pathogens in vegetable crops). Thus, the total cost of invading plant pathogens to crops in India is about $21.4 billion/yr.

4. Microbe Introductions Affecting Livestock

Several major diseases of livestock cause significant losses in India, most significantly foot-and-mouth disease. During 8 months in 1996, nearly 50,000 cases were reported at an estimated cost of $17,000/yr.

The combined cost of crop losses that can be attributed to nonindigenous plants, mammals, arthropods, and plant pathogens is $80 billion. Thus the negative effect of exotic species on Indian agriculture rivals that found in the United States.

VII. EXOTIC SPECIES INTRODUCTIONS IN BRAZIL

A. Nonindigenous Plants

Of the 55,000 plant species in Brazil, an estimated 21.1%, or 11,605, are alien species. Many of the alien species have become weeds and invaded a range of environments. In crop production, alien species make up 75% of the weed species. Weeds are estimated to destroy about 13.4% of Brazil's potential crop and pasture production, or about $12.3 billion/yr (see Table I).

B. Nonindigenous Vertebrates

1. Introduced Mammals

The number of introduced mammals in Brazil is estimated to be 30, including cats, dogs, sheep, goats, cattle, horses, pigs, and donkeys. The total number of mammal species in Brazil, including alien species, is 428.

On the outskirts of São Paulo in the community of Taboao de Sera, estimates suggest that there are 12,500 feral dogs and 4600 feral cats (Pimentel *et al.*, 2000b). The feral dogs attack livestock and other animals. Feral cats are also a serious problem, as they destroy birds and other native animals in Brazil.

It is estimated that there are about 320 million rats in Brazil. Assuming that each rat causes $10 in damages, the total cost of rats is estimated to be $3.2 billion/yr.

2. Introduced Birds

The number of native bird species in Brazil is 1635. Only three, the English house sparrow, common waxbill (*Estrilda astrild*), and common pigeon, are introduced species. The sparrow and pigeon populations cause major ecological and/or economic problems. The common waxbill is not considered a serious problem because it feeds primarily on the introduced guinea grass (*Panicum maximum*) and does little damage to this fast-growing grass (D. Pimentel, unpublished data).

C. Nonindigenous Invertebrates and Microbes

1. Introduced Arthropods

Invertebrate species number more than 100,000 in Brazil; about 70,000 are arthropod species. About 14.4% of potential crop production is destroyed by insects and mites, and approximately 35% of these pests are exotic. The calculated loss of crops to exotic insects and mites is estimated to be $4.6 billion/yr.

2. Introduced Crop Plant Pathogens

There are an estimated 100,000 species of microbes in Brazil; around 75% of the microbes that attack crops are exotic. Like most countries, most of these plant pathogens were introduced via alien crop species. Plant pathogens are estimated to cause 13.5% in crop losses each year. If 75% of the plant pathogen species are exotic, estimated losses from alien species total $18.3 billion/yr.

VIII. CONCLUSIONS

With approximately 400,000 nonindigenous species in various nations worldwide, if even a small fraction of these invaders are harmful, significant agricultural problems can result. While nearly all of our crop and livestock species are nonindigenous and have proven essential to the viability of the world's agriculture and economy, exotic species invasions do result in many negative financial effects. This article shows that the economic damages of these alien species to the agricultural economies of just six nations (including, in some cases, control costs) amount to approximately $247 billion each year. If this figure is extrapolated to all nations based on costs per person, the total cost of exotic species to world agriculture would be about $943 billion/yr.

Precise economic data for some of the most ecologically damaging biological invaders are not available. Mammals, including horses, water buffaloes, and goats, have been responsible for reducing the productivity of pastures and rangelands, as well as the extinction of many plant species in various ecosystems. In other areas, feral pigs and feral dogs have had serious impacts on agriculture, but as of yet few data are available concerning these animals.

Most of the exotic species in the world have been introduced within the past 70 years. Accelerated international trade and travel, plus rapidly increasing human numbers, guarantee that the threat from nonindigenous species invasions is still growing. The true challenge lies not in assigning precise dollar values to exotic species losses but in preventing further biological introductions and the resultant damage to managed and natural ecosystems. Although policies and practices to prevent the accidental and intentional introduction of exotic species are improving, the world's nations are woefully short of allocating sufficient resources to the problem in proportion to the risks. Both natural and managed ecosystems need to be protected from additional costly damages resulting from introduced species.

See Also the Following Articles

ENDANGERED ECOSYSTEMS • EXTINCTION, CAUSES OF • RANGE ECOLOGY • INTRODUCED SPECIES, EFFECT AND DISTRIBUTION • PESTICIDES, USE AND EFFECTS OF • PLANT INVASIONS

Bibliography

Babbitt, B. (1998). Statement by Secretary of the Interior on invasive alien species. Proceedings, National Weed Symposium, Bureau of Land Management Weed Page, April 8–10, 1998.

Bromilow, C. (1995). *Problem Plants of South Africa.* Briza Publications, Arcadia, South Africa.

Bryan, R. T. (1996). Alien species and emerging infectious diseases: Past lessons and future applications. In *Proceedings of the Norway/UN Conference on Alien Species, July 1–5, 1996* (G. T. Sandlund, P. J. Schel, and A. Viken, eds.), pp. 74–80. Norwegian Institute for Nature Research, Trondheim, Norway.

Crawley, M. J., P. H. Harvey, and A. Purvis. (1996). Comparative ecology of the native and alien floras of the British Isles. *Philos. Trans. Roy. Soc. London, Ser. B Biol. Sci.* 351, 1251–1259.

Emmerson, G., and J. McCulloch. (1994). *Feral Peril: Queensland's Introduced Plants and Animals.* Queensland Parliamentary Library, Brisbane, Australia.

Howarth, F. G. (1990). Hawaiian terrestrial arthropods: An overview. *Bishop Museum Occasional Papers* 30, 4–26.

Kurdila, J. (1995). The introduction of exotic species into the United States: There goes the neighborhood. *Environ. Affairs* 16, 95–118.

Layne, J. N. (1997). Nonindigenous mammals. In *Strangers in Paradise* (D Simberloff, D. C. Schmitz, and T. C. Brown, eds.), pp. 157–186. Island Press, Washington, D.C.

Lever, C. (1994). *Naturalized Animals: The Ecology of Successfully Introduced Species.* T. & A. D. Poyser Natural History, London.

Office of Technology Assessment (OTA). (1993). *Harmful Nonindigenous Species in the United States.* U.S. Government Printing Office, Washington, D.C.

Pimentel, D., L. Lach, R. Zuniga, and D. Morrison. (2000a). Environmental and economic costs associated with nonindigenous species in the United States. *BioScience,* 50(1), 53–65.

Pimentel, D., McNair, S., Janecka, J., Wightman, J., Simmons, C., O'Connel, C., Wong, E., Russel, L., Zern, J., Aquino, T., and Tsomondo, T. (2000b). Ecological and economic threat of alien plant, animal, and microbe invasions in the world. *Agriculture, Ecosystems and Environment, Submitted.*

Saxena, K. G. (1991). Biological invasions in the Indian subcontinent: Review of invasion plants. In *Ecology of Biological Invasions in the Tropics* (P. S. Ramakrishnan, ed.), pp. 53–73. International Scientific Publications, New Delhi.

Simberloff, D., D. C. Schmitz, and T. C. Brown. (1997). *Strangers in Paradise.* Island Press, Washington, D.C.

Smith, R. H. (1992). Rodents and birds as invaders of stored-grain ecosystems. In *Books in Soils, Plants, and the Environment: Stored-Grain Ecosystems* (D. S. Jayas, N. D. G. White, and W. E. Muir, eds.), pp. 289–323. Marcel Dekker, New York.

Temple, S. A. (1992). Exotic birds, a growing problem with no easy solution. *The Auk* 109, 395–397.

AGRICULTURE, INDUSTRIALIZED

Prabhu Pingali and Melinda Smale
International Maize and Wheat Improvement Center

GLOSSARY

alleles Variants of nucleotides within the genome that cause changes in a given protein.

crop genetic diversity The diversity of the sets of genes carried by different individuals within a crop species. It occurs in the form of nucleotide variation within the genome.

genetic diversity The diversity of the sets of genes carried by different organisms. It occurs in the form of nucleotide variation within the genome. When this variation causes a change in a given protein, the variants are called alleles. Allelic variation occurs at various genetic loci or gene positions within a chromosome.

genetic erosion The loss of genes from a gene pool attributed to the elimination of populations caused by factors such as the adoption of high-yielding varieties, farmers' increased integration into the market, land clearing, urbanization, and cultural change.

genetic variation The allelic variation that occurs at various genetic loci or gene positions within a chromosome. Genetically variable loci are termed polymorphic or are said to show polymorphism.

landraces These have a complex nature; therefore, it is not possible to give an all-embracing definition. Landraces are the originally adapted but variable crop populations on which farmers based their selections. They are geographically or ecologically distinctive populations which are conspicuously diverse in their genetic composition both between populations and within them. They have certain genetic integrity and are recognizable morphologically; farmers have names for them and different landraces are understood to differ in adaptation to soil type, time of seeding, date of maturity, height, nutritive value, use, and other properties.

loci Gene positions within a chromosome. Allelic variation occurs at various genetic loci.

spatial and/or temporal patterns of genetic variation The systematic changes in the alleles occurring at specific loci along spatial and/or temporal dimensions.

CROP GENETIC DIVERSITY IS THE BASIS of food production: Its pattern has changed significantly with the intensification and commercialization of agriculture. There is disagreement about the implications of these changes for the future. This article explores the nature of the links between the modernization of crop production and patterns of crop genetic diversity in the rice, maize, and wheat fields of the developing world.

I. INTRODUCTION

During the past 200 years, and in a more accelerated way since 1960, the pattern of crop genetic diversity in the fields of the developing world has changed fundamentally. The germplasm that dominates the area planted with major cereals has shifted from the locally adapted populations that farmers historically selected from seed they saved—often called "landraces"—to the more widely adapted seed types produced by scientific plant breeding programs and purchased by farmers—often called "modern varieties." These yield-enhancing seed types enabled the intensification of agriculture in areas of the world with high population densities. Initially, they diffused through the environments best suited for their production, spreading later—and unevenly—into less favored areas.

Economic growth, increasing incomes, and migration of labor from rural to urban areas lead to the commercialization of agriculture. Commercial crop production is characterized by the controlled application of water and other purchased inputs, such as fertilizer, fungicides, or pesticides to modern seed types. Sustained yield increases in commercial agriculture have depended not only on continued modifications in crop management practices but also on successive genetic improvements that are accomplished through recombining and exchanging diverse sources of germplasm. The spatial and temporal diversity among modern varieties in farmers' fields is determined in large part by the economic factors affecting their profitability and the performance of agricultural research institutions and seed industries.

In contrast, the genetic content and the geographical distribution of landrace populations are influenced more by natural selection pressures and the seed and crop management practices of traditional farming communities. Typically, it is not profitable to grow them with large amounts of purchased inputs since they are often grown principally for home consumption of food or feed and the size of their commercial market is limited. Seed markets for these varieties are generally local and rules for the exchange for seed, grain, and other products are more personal.

Today, the modernization of agriculture in the developing world is incomplete. In some areas with rain-fed, upland, or heterogeneous growing environments, landraces still dominate crop area or both modern varieties and landraces coexist. As the process of modernization continues, more farmers will be integrated into specialized product markets. Whereas others will work outside the farm and eventually leave it. Even then, local or export demand for certain quality attributes may provide market-based incentives for continued cultivation of landraces. In some places, there will be no alternative but to grow landraces for food since there may be no urban target for migration.

This article outlines some of the major implications of the modernization of agriculture for crop genetic diversity, focusing on the world's three principal cereals—rice, wheat, and maize. Wheat is a major tradable crop grown principally in temperate zones over a wide geographic area and often on relatively large, mechanized farms of industrialized countries. The world's main wheat-producing regions are southern Russia and the Ukraine, the central plains of the United States and adjacent areas in Canada, northwest Europe, the Mediterranean basin, southwestern Australia, and, in the developing world, India, China, and Argentina. Rice is the most important food grain in the developing world. Almost all the world's rice is both produced and consumed on numerous small farms within Asia, with less than 5% of production entering trade. India, China, and Indonesia are the major producers. Maize is the world's most widely grown cereal in terms of growing environments and, after wheat, the most extensively traded cereal. Although virtually all the wheat and rice traded internationally are destined for human consumption, most of the maize is used to feed livestock. Maize is a very important part of the human diet in Africa, parts of Latin America, and Asia. Major maize-producing regions include the United States and parts of Europe, Mexico, Brazil, Argentina, South Africa, India, and China.

The next section summarizes the factors driving the intensification of agriculture. Section III explains the implications of intensification for the patterns of genetic variation in the fields of the developing world. Section IV discusses how commercialization affects these patterns and presents hypotheses concerning the future for crop genetic diversity.

II. THE INTENSIFICATION OF CROP PRODUCTION

A. What Is the Intensification of Crop Production?

The intensification of agriculture refers to the increase in output per unit of land used in production, or land productivity. Population densities, expressed by the ratio of labor to land, explain much about where and under which conditions this process has occurred. The transition from low-yield, land-extensive cultivation systems to land-intensive, double- and triple-crop systems is only profitable in societies in which the supply of uncultivated land has been exhausted. The process of agricultural intensification has been observed in traditional and modern agricultural societies. The movement from forest and bush fallow systems of cultivation to annual and multi-crop cultivation systems, whereby plots of land are cultivated one or more times per year, has generally been influenced by increasing population densities in traditional societies. It is no accident that the modern seed-fertilizer revolution has been most successful in densely populated areas of the world, where traditional mechanisms for enhancing yields per unit area have been exhausted.

Intensive cultivation will also be observed in areas with lower population densities provided that soil conditions are suitable and markets are accessible. Intensification occurs in the less densely populated areas for two reasons: (i) Higher prices and elastic demand for output imply that the marginal utility of effort increases—hence, farmers in the region will begin cultivating larger areas, and (ii) higher returns to labor encourage migration into well-connected areas from neighboring regions with higher transport costs. Examples of regions with low population density but intensive, market-oriented production are the central plains of Thailand and parts of South America's southern cone. If the conditions described are not present, labor and other costs associated with intensive agriculture are substantially higher than its incremental economic returns.

Intensification of land use and the adoption of yield-enhancing technologies have occurred in traditional and modern agricultural systems. In the case of Africa, Pingali *et al.* (1997) documented the movement from shifting cultivation to permanent agriculture with increases in population densities and improvements in market infrastructure. As land became scarce, traditional farming communities across sub-Saharan Africa began to extract increasingly higher levels of output from their land through investments in land improvements and soil fertility management. The intensification of traditional farming systems is a process that the more densely populated regions of Asia had been through several decades and in some cases centuries earlier. The application of modern scientific innovations in the form of high-yielding seeds and fertilizer allowed the extremely land-scarce regions of Asia to achieve levels of land productivity that were not possible through the exclusive reliance on traditional systems of farmer innovation.

B. Seed Technical Change: Traditional and Modern

The genetic evolution of cultivated crops is closely interwoven with the evolution of human civilization and crop husbandry. The recognition of wild species of cereal crops was first accomplished by primitive societies of hunters and gatherers, who harvested them for food. The domestication of wild species occurred when societies of shifting cultivators first began to cultivate food crops. Sedentary or permanent agricultural systems led to the emergence of ancestors of modern-day landraces of cereal crops. Although landraces have evolved over thousands of years under farmer management of natural selection, varieties have been bred by plant breeding programs for about 100 years. The first high-yielding maize hybrids were developed about 50 years ago. The high-yielding varieties of rice and wheat with semidwarf stature were developed less than 50 years ago and have been successfully adopted only in intensive agricultural production systems. Modern cereal cultivars have developed through three main phases of selection: (i) subconscious selection by the earlier food growers in the process of harvesting and planting, (ii) deliberate selection among variable material by farmers living in settlements and communities, and (iii) purposeful selection by professional breeders using scientific methods.

The main attainment of the first phase was to make the crop more suitable for planting and harvest by humans, threshing or shelling, and consumption. Higher germination rates, more uniform growing periods, resistance to shattering, and palatability were some of the achievements of this effort. In the second phase, many farmers exerted pressures continuously in numerous directions, resulting in variable populations that were adapted to local growing conditions and consumption preferences. These are broadly known as landraces. During the third phase, fields of cereals have become more uniform in plant types with less spontaneous gene exchange. Planned gene migration has increased, however, with the worldwide exchange of germplasm. The products of the third phase are loosely referred to as "modern varieties."

The genetic changes embodied in seed can lead to change in the productivity of land (yield) both directly and indirectly, in combination with other inputs. A primary example of such technical change is the "green revolution." A term which has been used in various contexts to symbolize various types of social and technical change, it is first said to have been used by a USAID administrator to describe the rapid movement through Asia of new wheat and rice varieties coupled with increased use of fertilizer and irrigation. Here, we use it to refer specifically to the widespread adoption of semidwarf rice and wheat varieties in the developing world during the late 1960s and early 1970s. When grown with increased levels of fertilizer and a controlled water supply, these varieties performed significantly better than the varieties they replaced. Initially, they spread rapidly throughout many of the irrigated zones of the developing world where rice and wheat cultivation was concentrated and where population densities were high. Later, more widely adapted descendants of these varieties spread gradually into less favorable environments, including rain-fed areas with relatively modest production potential. Their diffusion was faster in the plains and valleys, diminishing up the hillsides and in more heterogeneous environments. The term modern varieties is also used to refer more exclusively to semidwarf varieties of rice and wheat.

Conservationists who observed the popularity of the green revolution varieties expressed concern for the possible loss of valuable genetic resources and narrowing in the genetic stock that serves as the basis for crop improvement (Frankel, 1970; Harlan, 1992; Hawkes, 1983). In the next section, the implications of the intensification of agriculture for crop genetic diversity are outlined. First, crop genetic diversity is defined. Some historical perspective is then provided on the spatial and temporal distributions of genetic variation in rice, wheat, and maize. Current estimates of areas planted with modern and landrace varieties of these cereals are reported. Evidence on the genetic diversity of modern varieties and landraces follows.

III. THE IMPLICATIONS OF INTENSIFICATION FOR CROP GENETIC DIVERSITY

Crop genetic diversity broadly defined refers to the genetic variation embodied in seed and expressed when challenged by the natural and human selection pressures that shape the environment in which it grows. In applied genetics, diversity refers to the variance among alternative forms of a gene (alleles) at individual gene positions on a chromosome (loci), among several loci, among individual plants in a population, or among populations (Brown *et al.,* 1990). Diversity can be measured by accessions of seed held in gene banks, lines or populations utilized in crop-breeding programs, or varieties cultivated by farmers (cultivars).

The relationship between precise quantitative measurements and what can be casually observed among the plants growing in a field, and between these measurements and what can be observed in other environments, is indirect. Crop genetic diversity cannot be literally or entirely observed at any point in time; it can only be indicated with reference to a specific crop population and analytical perspective. To understand the implications of agricultural intensification to crop genetic diversity, it is first necessary to gain a spatial and temporal perspective of the variation in crop plants.

A. Current Spatial Patterns of Genetic Variation

The adoption of modern cereal varieties has been most widespread in land-scarce environments and/or in areas well connected to domestic and international markets. Even in these areas, the profitability of modern variety adoption has been conditioned by the potential productivity of the land under cultivation. For instance, while modern rice and wheat varieties spread rapidly through the irrigated environments, their adoption has been less spectacular in the less favorable environments—the drought-prone and high-temperature environments for wheat and the drought- and flood-prone environments for rice. Maize, as discussed later, has an even spottier record in terms of farmer adoption of modern varieties and hybrids. For all three cereals, traditional landraces continue to be cultivated in the less favorable production environments throughout the developing world.

Recent estimates of the areas planted with modern varieties and landraces are shown in Table I. Based on data from a global wheat survey conducted by the Centro Internacional de Mejoramiento de Maiz y Trigo

TABLE I

Percentage Distribution of Rice, Wheat, and Maize Area by Type of Germplasm in the 1990s

	Wheat			Maize			Rice		
Region	Semidwarf improved	Tall improved	Landraces/ unknown	Hybrid	Improved open-pollinated	Landraces/ unknown	Semidwarf improved	Other improved	Landraces/ unknown
Sub-Saharan Africa[a]	66	14	20	38	8	54	25	15	60
West Asia/North Africa	66	10	24	22	7	71	11		
Asia	86	7	6	70	7	23	73	13	14
Latin America	90	9	1	43	5	52	59	36	5
All developing countries	81	9	11	53	7	40	71		
Industrialized countries	55	45	Trace	99	1	Trace	78		

Note. Sources: IRRI World Rice Statistics (1995, 1999), CIMMYT Global Maize and Wheat Impacts Surveys (1992, 1997), Heisey et al. (1999) Morris and López-Pereira (1999), Luis Roberto Sanint (CIAT), and Timothy Dalton (WARDA).

[a] Data for rice in sub-Saharan Africa are West Africa only.

(CIMMYT) in 1997, approximately 80% of the wheat area in the developing world was sown to semidwarf varieties, with the remainder split almost equally between improved tall varieties and landraces or varieties with unknown ancestry. The relative importance of tall wheat varieties remains greater in the industrialized than in the developing countries, probably for reasons related to wheat-growing environment and management practices.

Most of the extensive area sown to wheat landraces is found in parts of Turkey, Iran, Afghanistan, and Ethiopia. Pockets of landrace diversity for special traits may also be found throughout the Mediterranean region (Morocco, Tunisia, Syria, Egypt, Cyprus, Portugal, Spain, and Italy) as well as Nepal, Outer Mongolia, and Tibet, although field research and data analysis would be needed to confirm these hypotheses. Work in Mexico, Guatemala, Honduras, Peru, and Bolivia indicates the presence of unique landrace varieties that were probably introduced by Spanish immigrants in the sixteenth century. No significant areas are believed to be sown to wheat landraces in the former Soviet Union or in Eastern Europe, despite the historical importance of these areas for major progenitors such as the so-called "Turkey" and "fife" wheat. The loss of landrace populations in the former Soviet Union was no doubt a consequence of collective agriculture that was practiced in the region for several decades.

Approximately three-fourths of the rice area in Asia, which produces most of the world's rice, is thought to be sown to semidwarf varieties. Semidwarf varieties dominate the irrigated rice ecosystems of Asia and cover substantial areas in the favorable rain-fed lowlands (M. Jackson and G. Khush, personal communication). In the more marginal rain-fed lowland environments, the deep-water environments, and the upland rice environments, farmers continue to grow landraces adapted to those particular environmental niches. Although more accurate data concerning rice landraces in Asia are now being compiled at the International Rice Research Institute (IRRI), the most comprehensive data from a national-level sample survey conducted by the Department of Agricultural Extension in Bangladesh (1996–1997) suggests that although farmers in that country still grow large numbers of named varieties including landraces (more than 300), more than 20 modern varieties cover nearly 65% of the total rice area. In West Africa, semidwarf varieties cover essentially all the irrigated lowlands and one-third of the rain-fed lowlands, but landraces and other improved varieties appear to occupy most of the area in uplands, mangrove, and deep-water flooded areas. Particularly in the deep-water/floating rice environment near Mopti in Mali, Birnin Kebbi in Nigeria, and in northern Guinea are found *Oryza glaberrima* (African rice) and tall traditional *O. sativa* (Asian rice) cultivars. The West African Rice Development Association (WARDA) has recently developed hybrid crosses of *O. sativa* and *O. glaberrima* species. In East Africa, the area in Madagascar seems to be approximately evenly split between traditional and modern varieties. In Latin America, tall and landrace varieties cover most of the upland area, whereas few are found in the irrigated and rain-fed lowlands.

In the uplands of Asia, traditional varieties still dominate, with "pockets" of important diversity in the Bastar Plateau (Madhya Pradesh, eastern India), parts of northern Bangladesh and the Chittagong Hill Tracts in that country, northern states of Myanmar, almost all of Laos, and parts of Cambodia (M. Jackson and G. Khush, personal communication). Oka (1988) reported that the genetic diversity in landraces of Asian rice was most prevalent in the area extending over Assam, Bangladesh, Burma (now Myanmar), Thailand, Laos, and Yunnan, China. Although Oka also reported that diversity was relatively high in Indonesia, which lies outside the center of domestication, much of Indonesian ricelands are now planted with modern varieties.

Data from a global maize surveys conducted by CIMMYT in 1992 and 1997 indicate that, relative to wheat and rice, far less of the maize area in the developing world is planted with maize types released by plant breeding programs. Hybrids appear to occupy an increasing proportion of the area planted with modern maize types, but in zones such as Latin America, most of the maize area is still planted with landraces.

At least some of the area listed under landraces is planted with populations that result from the genetic integration of modern varieties with landraces when farmers save seed or plant seed of different types in adjacent fields. Similarly, a substantial proportion of the maize area in sub-Saharan Africa is planted with advanced generations of improved varieties whose seed farmers could not afford to replace on a regular basis introgressed with landrace populations brought to the continent with the slave trade several centuries ago. Although there appears to be a high proportion of maize landraces grown in West Asia and North Africa, the total area in maize is limited in this region and many of these are likely of unknown origin. In the industrialized world, a negligible percentage of maize area is planted with either improved open-pollinated varieties or landraces. These are specialty maizes or "heirloom" varieties grown for fresh consumption, popcorns, or ornamental corns such as those marketed on holidays in the United States.

The data in Table I confirm that although much of the genetic variation in the world's wheat and rice area is shaped today by the efforts of modern plant breeders, a substantial part of the genetic variation in the maize grown in the developing world remains in the hands of farmers, some of which are among the poorest in the world. In some zones of production, such as Turkey and Iran for wheat or parts of Latin America for maize, fairly large contiguous areas may be planted to landraces. Often, in regions of crop domestication and diversity, landraces persist "as patches and islands of farming systems" (Brush, 1995, p. 246). Harlan (1992, p. 147) invoked the term "microcenters of diversity" to describe "relatively small regions, 100–500 km across, in which may be packed an astonishing variation" within extensive areas of apparent uniformity in parts of Turkey, the Caucasus, Iran, and Afghanistan. Oka (1988) called the Jeypore Tract in India a microcenter of diversity in Asian rice. Jackson (personal communication) refers to some of the upland rice areas of Asia as "pockets" of diversity.

The change in the crop genetic landscape from predominantly traditional to largely modern patterns of genetic variation occurred during the past 200 years and at an accelerated rate since the 1960s. Whether the change to modern varieties has resulted in a narrowing of genetic diversity remains largely unresolved for many reasons, which are discussed in the following sections and further in Smale (1998) and Wood and Lenné (1997).

B. Comparing Genetic Diversity in Landraces and Modern Varieties

A major problem in assessing whether genetic narrowing has occurred with the replacement of landraces by modern varieties is the magnitude of the sampling and measurement effort that would be required to test the hypothesis in a meaningful way. In most cases it would be impossible to locate reliable samples of the landraces originally grown in an area now planted with modern varieties since this process occurred over time and unevenly across environments.

In some sense, the genetic diversity of landraces and modern varieties is incomparable by definition since its structure is distinct for each. Hawkes (1983) wrote that landraces, which are mixtures of genotypes, "could not even be called varieties." He called the range of genetically different varieties available to breeders "the other kind of diversity" (pp. 100–101). Harlan (1992) described a landrace as "an integrated unit" of "component genotypes" that have adjusted to one another over the generations as well as to the local environment, both natural and man-made (p. 148). A field planted with a rice or wheat landrace may be viewed by a plant breeder as a mixture of several lines but viewed by the farmer as one single variety because of its recognizably distinct agronomic, processing, or consumption characteristics. Vaughan and Chang (1992) described traditional rice types that are mixtures or composites of morphologically distinguishable types grown together deliberately.

The breeding system of the crop plant also affects the structure of genetic variation. Landraces generally contain some heterozygous material, though the extent of segregation is considerably greater for open-pollinated than for self-pollinated species. Maize is a cross-pollinating species, and maize in Mexico is often cited as an example of deliberate manipulation of the composition of landraces by farmers. Landrace varieties of maize evolve continuously through the purposeful mixing of seed lots of the same varieties or introgression by farmers, as well as inadvertently when fields of maize planted with different varieties flower simultaneously. A single "race" or race complex, as understood by maize geneticists, contains numerous genetically distinct farmers' varieties.

The distinction between modern varieties and landraces can also be blurred for predominantly outcrossing crops such as maize, making it difficult to determine which population is under study. Many small-scale, subsistence-oriented maize farmers promote hybridization between improved varieties and landraces by growing them together or in neighboring fields and producing what farmers in Mexico call "creolized" varieties. Also termed "rustication" or simply "adaptation," this process may enable improved varieties to fit better the need of local farmers.

These points imply that comparing counts of landraces and modern varieties as an index of genetic narrowing may not make sense. They also imply that even if reliable samples of the landraces originally cultivated in an area could be obtained, analyses comparing their genetic diversity might provide only part of the answer regarding genetic narrowing. Although the landrace in the farmers' field is a heterogeneous population of plants, it is derived from generations of selection by local farmers and is therefore likely to be local in adaptation. The plants of a modern variety are uniform but the diverse germplasm in the genetic background may enable them to adapt more widely. The diversity in a modern variety may not be expressed until challenged by the environment. On the other hand, the landrace may carry an allele that occurs rarely among modern varieties and is a potentially valuable source of genetic material not only for the farmer that grows it today but also for future generations of producers and consumers.

C. When Did Genetic Narrowing Occur?

Another problem in assessing the relationship of modern varieties to genetic narrowing is the temporal point of reference. Porceddu *et al.* (1988) described two major stages of genetic narrowing in wheat during modern times. The first occurred in the nineteenth century when scientific plant breeding responded to the demand for new plant types. Farming systems emerged that were based on the intensive use of land and labor, livestock production, and the use of organic manure. Changes in cultivation methods favored genotypes that diverted large amounts of photosynthates into the ear and grain. Bell (1987) reports that the engineering innovations of the late nineteenth century led to the establishment of extensive wheat-growing areas in North America, Australia, and parts of South America. Mechanization of agriculture dictated uniformity in plant type.

According to Porceddu *et al.* (1988), a second stage of narrowing occurred in the twentieth century, when genes were introduced to produce major changes in plant type. Use of the dwarfing genes *Rht1* and *Rht2*, for example, conferred a positive genotype-by-environment interaction in which yield increases proved greater given a certain combination of soil moisture, soil fertility, and weed control. Varieties carrying these dwarfing genes were developed by N. Borlaug with the national breeding program in Mexico and later by the International Maize and Wheat Improvement Center (CIMMYT). They became known as the green revolution wheats.

Evenson and Gollin's (1997) summary of the history of rice breeding suggests a process of continual expansion and narrowing of the genetic pool. Organized breeding efforts probably date earlier than 1000 AD in China. Modern efforts can be traced to the late nineteenth century in several parts of Asia. In temperate east Asia, the first significant advances were made by Japanese farmers and scientists when they developed relatively short-statured and fertilizer-responsive cultivars. Known as the *rono* varieties, these belonged to the *japonica* class of rice and were widely cultivated in Japan as early as the 1890s. During the Japanese occupation of Taiwan in the early part of the twentieth century, Japanese scientists sought to adapt these varieties to the more tropical conditions of Taiwan. At the same time, researchers in tropical Asia were seeking more productive varieties of rice from the *indica* and *javanica* classes of rice. After World War II, the United Nations Food and Agricultural Organization initiated a program to cross *indica* rice with *japonicas* as a means of increasing rice yields, culminating in the formation of the IRRI and the green revolution varieties of rice.

To Vaughan and Chang (1992), genetic narrowing in modern rice began early in this century. Development projects, population increases, and forest clearing in Asia were the primary causes of the loss of wild and cultivated rice landraces. In the Mekong Delta, the replacement of traditional deep-water rice by irrigated rice occurred with drainage and irrigation schemes that were introduced during the French colonial period.

Goodman (1995) reports that the major portion of the variability now found in maize developed before European contact (circa 1500), and several of the most widely grown races, including the commercially important

Corn Belt dents, developed later. During the "corn show era" in the nineteenth century, U.S. farmers exhibited their open-pollinated varieties locally and emphasis was placed on uniformity and conformity to an "ideal type." By the early 1950s, essentially all of the maize grown in the Corn Belt was double-cross hybrid. After the late 1950s, increasingly more farmers in the U.S. Corn Belt grew single-cross rather than double-cross hybrids. Because single-cross seed must be produced on an inbred line, this type of selection contributed to a marked loss of variability in U.S. breeding materials. According to Goodman, a countervailing influence during the past 25 years has been the emphasis by public researchers on development of improved maize populations.

Not all scientists agree about what constitutes genetic narrowing or precisely when such narrowing has occurred. For instance, in contradiction with Porceddu *et al.* (1988), Hawkes (1983) cites the introduction of *Rht1* and *Rht2* genes into Western wheat breeding lines as an example of how diversity has been broadened by scientific plant breeders. The Japanese line Norin 10 carried the dwarfing genes from the landrace Daruma, believed to be of Korean origin. Similarly, the efforts to increase rice yields by crossing *japonica* and *indica* classes of rice extended the gene pool accessible to rice breeders. As these examples suggest, in modern agriculture, today's broadening of the genetic pool in a plant breeding program may lead to a narrowing of the breadth of materials grown by farmers precisely because such innovations often produce varieties that are popular.

D. Trends in Genetic Diversity of Modern Varieties of Rice, Wheat, and Maize

Part of the concern for genetic narrowing is based on the perception that, with time, conventional plant breeding practices inevitably restrict the genetic base of modern varieties. The evidence from studies on the parentage of modern varieties lends little support to the view (Witcombe, 1999). In an analysis of genealogies of 1709 modern rice varieties, Evenson and Gollin (1997) found that although a variety released in the 1960s had 3 landraces in its pedigree, recent releases have 25 or more. The complexity of rice pedigrees, in terms of parental combinations, geographical origin, and number of ancestors, has expanded over time. A similar pattern has been shown for about 800 wheat varieties released in the developing world since the 1960s (Smale, 1997). The average number of distinct landraces found in bread wheat pedigrees increased from approximately 20 in the mid-1960s to about 50 in 1990.

Skovmand and de Lacy (1999) analyzed the distance among coefficients of parentage for a historical set of CIMMYT wheat varieties during the past four decades. Their results show a rate of increase in genealogical diversity that is positive but decreases over time, with marked expansion in genealogies from 1950 to 1967 and gradual flattening through the 1990s. If progenitors were recycled and reused, the distance among them would decrease over time and the slope of the line would be negative.

Less evidence is available worldwide on trends in the pedigrees or ancestry of maize varieties than for rice and wheat, in part because this information is confidential in an increasingly privatized industry. Following the epidemic of corn blight in the U.S. crop in 1970, the National Research Council (1972) concluded that the genetic base of maize in the United States was sufficiently narrow to justify concern. Duvick (1984) found that during the 10 years following the 1970 epidemic, breeders had broadened their germplasm pools.

Molecular markers, like genealogies, can be used to construct indicators of the latent diversity in a set of crop populations. Using molecular markers, Donini *et al.* (2000) concluded that there is no objective evidence to support the assertion that modern plant breeding has reduced the genetic diversity of U.K. wheats since 1930. Recent molecular evidence for a set of CIMMYT wheats indicates that genetic distance has been maintained among major parents and popular varieties during the past 30 years (unpublished data). Since many of the varieties of spring bread wheat grown in the developing world have a combination of CIMMYT and locally bred materials in their ancestry (Heisey *et al.*, 1999), these data represent a lower bound on actual genetic diversity. Furthermore, the genetic diversity that is accessible to conventional plant breeders today includes not only spring bread wheat, of course, but also wheat types with different growing habits, close relatives, and wild grasses. Techniques of biotechnology may traverse the species barriers faced by conventional breeders.

IV. THE COMMERCIALIZATION OF CROP PRODUCTION

A. What Is Commercialization of Crop Production?

Economic growth, urbanization, and the withdrawal of labor from the agricultural sector have led to the increasing commercialization of agricultural systems.

Subsistence-oriented monoculture food production systems give way to a diversified market-oriented production system. Agricultural commercialization means more than the marketing of agricultural output: It means that product choice and input use decisions are based on the principles of profit maximization. Commercial reorientation of agricultural production occurs for the primary staple cereals and for the so-called high-value cash crops. Commercialization of agricultural systems leads to greater market orientation of farm production; progressive substitution of nontraded inputs in favor of purchased inputs; and the gradual decline of integrated farming systems and their replacement by specialized enterprises for crop, livestock, poultry, and aquaculture products (Pingali, 1997).

On the demand side the process of agricultural commercialization is triggered by rapid income growth and the consequent diversification in food demand patterns. A slowdown in income-induced demand for rice and for coarse grains is accompanied by a shift of diets to bread and higher valued foods such as meat, fruit, and vegetables. These dietary transitions are induced by the growth in per capita income and by the rapid migration of population to urban areas. The need to provision the rapidly growing cities of the world also acts as an impetus for the transformation of food production systems.

On the supply side, growing factor scarcities contribute to the demise of subsistence agricultural systems. Although growing land and water scarcity can be compensated for with increasing scientific knowledge and farmer management, farmer time required for sustaining productivity and profitability of intensive food production systems will become increasingly scarce. The collapse of subsistence systems will come about because of the competing demands for farmers' time. Although the speed of the structural transformation differs substantially across countries they are all moving in the same direction.

B. Seed Industries

As the orientation of crop production shifts from subsistence toward commercial objectives, the locus of crop improvement and seed distribution moves from individual farmers toward an organized seed industry composed of specialized private and public organizations. In terms of an increased reliance on commercially produced seed, this has occurred substantially faster for maize than for rice and wheat. In a stylized depiction of the maize seed industries in developing countries, subsistence production is characterized by open-pollinated varieties improved through farmer selection and on-farm seed production with local seed markets governed by custom. In a fully commercial system, the predominant seed type is a hybrid that is purchased annually. Seed is traded globally and is a product of specialized research that is both privately and publicly funded. The exchange of seed and the genetic resources used to improve it are enabled and protected by strict forms of intellectual property rights.

For rice and wheat, which are self-pollinating crops, the incentives for privatization of research have not always been as strong as those for maize, although this depends on the institutional and economic context. In industrialized countries, profound changes in science and in intellectual property protection during the past 20–30 years have been associated with a higher rate of investment in agriculture by the private sector than the public sector and a shift in the composition of private investment from agricultural machinery and processing into chemical research and plant breeding. Although privatization is greatest in the maize seed industry, it is also occurring in wheat to a lesser extent and particularly in Europe. There is very little private sector rice breeding anywhere in the world. Almost all the research on rice has been conducted by the public sector, and most of this has taken place in Asia. In the developing world, there is increasing privatization of the maize seed industry but rice and wheat remain primarily public.

In commercial systems of rice, wheat, and maize, recent changes in the structure of the seed industry are likely to have implications for the utilization of modern patterns of genetic variation. The global seed industry has integrated both vertically (within production processes) and horizontally (among production processes) into "life science" firms that combine seed, chemical, and pharmaceutical businesses. As part of this structural change, firms are engaging in strategies to ensure more exclusive proprietary rights, including, for the first time in history, patents on genetically modified organisms.

The implications of these changes for the exchange and utilization of the genetic resources that are used in breeding modern varieties are unknown. Efforts are under way to harmonize intellectual property regimes globally through international trade agreements, but differences between developed and developing countries, as well as among developing and developed countries, pose challenges. Small public plant breeding programs in developing countries are not on the same footing with respect to investments and legal clout as the life science conglomerates. The maize, wheat, and rice industries are likely to be affected in different ways given the nature of economic incentives associated with seed reproduction. Patents are only one type of intellectual property right; in addition to intellectual property rights, national seed regulatory systems will have a strong impact on farmers' access to seed and the conservation of plant genetic diversity.

V. THE IMPLICATIONS OF COMMERCIALIZATION FOR CROP GENETIC DIVERSITY

Agricultural commercialization influences the extent of crop genetic diversity in two ways: (i) through changes in land use patterns and (ii) through crop choice changes in the irrigated as well as the rain-fed environments. The organization and management of food production systems in both the irrigated and the rain-fed environments are affected by economic growth. The opportunity cost of family labor can be expected to increase equally in the high- and low-potential areas since the populations in both environments are responding to nonagricultural employment opportunities. The declining viability of subsistence production systems can also be expected to be similar. The movement from subsistence to market-oriented, rain-fed production could follow a general pattern: (i) the abandonment of highly drought-prone environments, especially in areas where the opportunities for groundwater exploitation are limited; (ii) the shift from small subsistence farms to mechanized cultivation of large farms; and (iii) where dry season water supplies are available, increased areas under vegetables, feed grain and fodder crops, and other high-valued crops. Cereal crop production would continue to have a comparative advantage in the rain-fed environment, primarily because of the high cost of modifying the environment in order to make it suitable for noncereal crops. Low-input, low-yield cereal production systems, rice and maize, grown on consolidated holdings may emerge as the most viable option for the rain-fed environments. The irrigated environments, on the other hand, would shift from being predominantly under cereal monoculture to a highly diversified production system. The implications of these changes in crop choice on genetic diversity in the irrigated and rain-fed environments are discussed in the following sections.

A. Implications When Intensification Has Occurred

In the irrigated rice and wheat production zones of the developing world, commercialization has had little impact on crop genetic diversity beyond that of agricultural intensification. Since the modern varieties grown there are varieties rather than hybrids, their seed is saved and spreads now, as in the past, from farmer to farmer. This will change if hybrids, or certain types of transgenic varieties that require annual seed purchase, are developed. It may also change if the demand for labor-saving technologies such as herbicides leads to the use of herbicide-tolerant varieties. These must be purchased annually to prevent the carryover of weed seed.

The effects of commercialization on maize crop genetic diversity, independent of those associated with agricultural intensification, are much more pronounced because of the historical importance of hybrids relative to improved, open-pollinated varieties. The distinctive biological properties of maize plants (in particular, their propensity for open pollination and their tendency to segregate) make it difficult for farmers to maintain the genetic purity of maize seed saved from their own harvest. Commercial maize growers are therefore dependent on reliable external sources of affordable seed in a way that growers of self-pollinated rice and wheat are not. The reliance on the seed industry will continue to grow for maize if farmer use of genetically engineered seed increases in importance in the future.

The survival of landrace diversity for cereal crops in the high-potential, irrigated environments would depend on farmer incentives for maintaining that diversity. To a large extent, farmer incentives to do so would depend on the market demand for the unique quality characteristics that are present in some of the landraces. The importance of Basmati rice in the irrigated production zones of India and Pakistan is an example of how market demand for quality can influence the survival of traditional varieties and landraces. Even where modern varieties are used exclusively, diversity within the plant has increased over time, as discussed previously, by the introduction of new gene pools through breeding.

B. Implications When Intensification Has Not Occurred

The areas of the developing world where modern varieties are not widely grown are typically marginal for production of the crop or are inadequately served by markets and infrastructure. In some areas, agricultural research has been unable to produce varieties demonstrating an obvious yield advantage or commercial seed systems have not had the incentive to do so because of the small size of the market or fluctuating effective demand.

As suggested previously, this may remain the case for sizable portions of the developing world's maize area. The proportion of cultivated area that is irrigated is far less for maize than for rice and wheat, whereas the use of purchased inputs remains modest. Most of the farmers who grow maize in developing countries face difficult and variable maize production environments and cultivate it with the primary objective of meeting subsistence requirements. These farmers have little incentive to make investments in fertilizer, pesticides, and other modern means of coping with disease and weather since their traditional varieties do not respond as well to these as modern varieties. In many of these production zones, it is not easy to breed well-adapted materials and there are few profits to be earned for seed companies.

In the most difficult environments, commercialization is likely to lead to the complete abandonment of crop production, as has already occurred in parts of Asia. When there are limited opportunities for migration but environments are too marginal for specialized agricultural production to be profitable, farmers may remain on small landholdings and grow landraces for subsistence.

Even when a zone may be suitable for the production of modern varieties, the development of commercial seed systems is not sufficient to ensure that they will replace landraces in the near future because markets are imperfect. In some local communities, the specific varietal traits demanded by farmers (grain quality, fodder, and suitability for a certain soil type) cannot be obtained through the production of modern varieties or procured through impersonal market transactions so that farmers must rely on their own production or that of nearby farmers for their supply of a valued attribute. The specialized uses of certain landrace varieties for medicinal purposes, rituals and festivals, and culinary practices have been extensively documented.

Small-scale farmers' choice to grow more than one variety simultaneously is likely to reflect their need to address numerous concerns that no single variety can satisfy (Bellon, 1996). Farmers often choose to grow both landraces and modern varieties. Zimmerer (1999) found that the capacity of farmers to grow diverse food plants (including maize) in Peru and Bolivia depends on whether they can cultivate them in combination with commercially developed, high-yielding varieties. Vaughan and Chang (1992) noted that the rapid changes that occur with natural calamities are more likely to have a greater impact on the loss of rice genetic diversity than the farmer-driven, incremental changes that are regularly occurring, many of which enhance diversity (Dennis, 1987). Meng *et al.* (1998) concluded that multiple factors, including missing markets, yield risk, grain quality, and agroclimatic constraints, influence the probability that a Turkish household will grow a wheat landrace; a change in any single economic factor is unlikely to cause farmers to cease growing it.

Viewed in the conventional microeconomic literature as partial adoption, this observed pattern has been explained theoretically through attitudes toward risk and uncertainty, nonexisting markets, and differential soil quality or nutrient response combined with fixity or rationing (Meng, 1997; Smale *et al.*, 1997). Although treated as a transitional period to full adoption (or replacement), the coexistence of modern varieties and landraces may represent an equilibrium if one or several of these aspects persist despite economic change. Then there are locally based economic incentives for farmers to continue to grow landraces.

Even when the pressures for market integration are strong, the coexistence of modern varieties and landraces may also persist with certain types of market-based incentives, as discussed previously. In the early phases of economic growth when rural populations move to urban areas, market integration exerts pressures for uniformity in the attributes of coarse grains. Localized preferences diminish in favor of cheaper, bulk-marketed grains. The elasticity of demand for staple grains declines as income increases, and it is sometimes negative. Generally, rice substitutes for maize, and wheat substitutes for rice.

The income elasticity of demand for attributes of the grains may be higher, however, than the income elasticity of demand for the cereal (Pingali *et al.*, 1997). For example, a notable pattern or rice consumption is that, with growing incomes, people express preferences for higher quality rice once their calorie needs have been met. High-income consumers spend more on rice by paying higher prices for varieties with preferred eating quality which they substitute for the lower quality variety consumed when the their income levels were lower. In Asia, traditional varieties are generally of higher quality and fetch premium prices in the market. Thailand still grows low-yielding traditional rain-fed varieties extensively for the export market. When the income level was low, South Korea used to grow the modern "tongil" variety, but this was replaced by relatively low-yielding traditional *japonica* rice as consumers expressed preference for *japonicas* by offering higher prices. In response to these market signals farmers are eager to grow even low-yielding, high-quality rices because the higher prices more than compensate for their lower yields. Because rice scientists have had limited success in developing high-yielding cultivars with better eating quality, the price difference between the standard- and high-quality varieties has been increasing in Asian markets.

The post-industrial agricultural economy is characterized by growth in demand for an array of increasingly specialized goods and services. Some product quality attributes are associated with features of the production process (organic/inorganic). Some are extrinsic (origin or effects on animal welfare), whereas others cannot be discerned without laboratory tests (genetically modified organisms). The elasticity of demand for such attributes is likely to increase with very high levels of income. Under these conditions, global market integration may provide market-based incentives for continued cultivation of patches of diverse landraces. In addition to their demonstrated private economic value to the farmers who grow them, some of these hold potential for niche markets and exports.

VI. CONCLUSIONS

In areas of the world that are more favorable for agricultural intensification, the conversion from landrace varieties to modern varieties has been almost complete for rice, wheat, and maize. In some of the remaining crop-growing environments, and especially in relatively small pockets of crop diversity called micro-centers, farmers who are linked to commercial agriculture through labor markets or specialized product markets continue to grow landraces—often in combination with modern varieties. Some cannot obtain or afford to purchase seed on a routine basis, so they both purchase and save the seed of modern types. Others purposefully adapt modern varieties to their own conditions by saving and selecting the seed, genetically integrating modern and landrace types.

The structure of genetic diversity is distinct for modern varieties and landraces. Both are essential to the future food supply. Conservationists propose that landrace diversity must be maintained not only in preserved stocks called *ex situ* collections but also *in situ*, or in the fields of farmers. The future of landrace cultivation appears to be uncertain. Although some view the replacement of landraces by modern varieties as an inevitable product of agricultural commercialization, idiosyncratic growing environments and consumer preferences may provide economic incentives for their continued cultivation by farmers—although on a limited scale. It appears unlikely that modern varieties of rice, wheat, and maize will entirely replace landraces in the near future, although it is difficult to postulate about equilibrium areas planted with each type of germplasm since the equilibrium itself shifts with technical and economic change.

Continued genetic improvement does not necessarily lead to loss of genetic diversity in areas where modern varieties dominate, especially when access to germplasm is relatively unrestricted and innovative plant breeding strategies may be employed. Access to diverse sources of germplasm is therefore of great importance to the success of public and private breeding programs for the supply of varieties in modern agriculture. The continued advances in yield potential that are a necessary (although not a sufficient) condition for alleviating hunger are thought to depend on increasingly complex combinations of genes and novel alleles. Landraces and wild relatives have served as repositories for resistance to biotic and abiotic stress when these are absent in advanced breeding materials. Even in the parts of the world where the "ancient patterns of diversity" may still be found, access to the products of modern plant breeding is often integrated economically or genetically to generate more resilient and sustainable systems.

See Also the Following Articles

AGRICULTURE, SUSTAINABLE • AGRICULTURE, TRADITIONAL • CROP IMPROVEMENT AND BIODIVERSITY • GENETIC DIVERSITY • HERBICIDES • PESTICIDES, USE AND EFFECTS OF

Bibliography

Bell, G. D. H. (1987). The history of wheat cultivation. In *Wheat Breeding: Its Scientific Basis* (F. G. H. Lupton, Ed.). Chapman & Hall, London.

Bellon, M. R. (1996). The dynamics of crop infraspecific diversity: A conceptual framework at the farmer level. *Econ. Bot.* **50**, 26–39.

Brown, A. H. D., Clegg, M. T., Kahler, A. L., and Weir, B. S. (Eds.) (1990). *Plant Population Genetics: Breeding and Genetic Resources*. Sinauer, Sunderland, MA.

Brush, S. B. (1995). *In situ* conservation of landraces in centers of crop diversity. *Crop Sci.* **35**, 346–354.

CIMMYT Economics Program (1999). Global Maize and Wheat Impacts Surveys, 1992, 1997. CIMMYT: Mexico, D.F.

Dennis, J. V. (1987). Farmer management of rice variety diversity in Northern Thailand. Ph.D. thesis, Cornell University, Ithaca, NY.

Donini, P., Law, J. R., Koebner, R. M. D., Reeves, J. C., and Cooke, R. J. (2000). Temporal trends in the diversity of UK wheat. *Theor. Appl. Genet.* (in press).

Duvick, D. N. (1984). Genetic diversity in major farm crops on the farm and in reserve. *Econ. Bot.* **38(2)**, 161–178.

Evenson, R. E., and Gollin, D. (1997). Genetic resources, international organizations, and improvement in rice varieties. *Econ. Dev. Cultural Change* **45(3)**, 471–500.

Frankel, O. H. (1970). Genetic dangers of the Green Revolution. *World Agric.* **19**, 9–14.

Goodman, M. M. (1995). Maize. In *The Evolution of Crop Plants* (J. Smartt and N. W. Simmonds, Eds.), pp. 193–202. Wiley, New York.

Harlan, J. R. (1992). *Crops and Man*. American Society of Agronomy/Crop Science Society of America, Madison, WI.

Hawkes, J. G. (1983). *The Diversity of Crop Plants*. Harvard Univ. Press, Cambridge, MA.

Heisey, P. W., Lantican, M., and Dubin, H. J. (1999). Assessing the benefits of international wheat breeding research: An overview of the global wheat impacts study. In *CIMMYT 1998–99 World Wheat Facts and Trends. Global Wheat Research in a Changing World: Challenges and Achievements* (P. L. Pingal, Ed.). CIMMYT, Mexico City, Mexico.

IRRI (1995). World Rice Statistics, 1993–1994. IRRI: Manila, Philippines.

IRRI (1999). World Rice Statistics, 1998–1999. IRRI: Manila, Philippines.

Meng, E. C. H. (1997). Land allocation decisions and in situ conservation of crop genetic resources: The case of wheat landraces in Turkey. Ph.D. thesis, University of California, Davis.

Meng, E. C. H., Taylor, J. E., and Brush, S. B. (1998). Implications for the conservation of wheat landraces in Turkey from a household model of varietal choice. In *Farmers, Gene Banks and Crop Breeding: Economic Analyses of Diversity in Wheat, Maize, and Rice* (M. Smale, Ed.), pp. 127–143. Kluwer/CIMMYT, Boston.

Morris, M. L., and López-Pereira, M. A. (1999). *Impacts of Maize Breeding Research in Latin America, 1966–67*. CIMMYT, Mexico City, Mexico.

National Research Council (1972). Genetic Vulnerability of Major Crops. National Academy of Sciences, Washington, D.C.

Oka, H. I. (1988). *Origin of Cultivated Rice*. Japan Scientific Societies Press/Elsevier, Tokyo.

Pingali, P. L. (1997). From subsistence to commercial production systems: The transformation of Asian agriculture. *Am. J. Agric. Econ.* **79**, 628–634.

Pingali, P. L., Hossain, M., and Gerpacio, R. V. (1997). *Asian Rice Bowls: The Returning Crisis*? CAB International, Wallingford, UK.

Porceddu, E., Ceoloni, C., Lafiandra, D., Tanzarella, O. A., and Scarascia Mugnozza, G. T. (1988). Genetic resources and plant breeding: Problems and prospects. In *The Plant Breeding International, Cambridge Special Lecture*, pp. 7–21. Institute of Plant Science Research, Cambridge, UK.

Skovmand, B., and de Lacy, I. H. (1999). Parentage of a historical set of CIMMYT wheats. 1999 annual meeting abstracts. American Society of Agronomy, Madison, WI.

Smale, M. (Ed.) (1998). *Farmers, Gene Banks and Crop Breeding Economic Analyses of Diversity in Wheat, Maize, and Rice*. CIMMYT/Kluwer, Mexico/Norwell, MA.

Smale, M., Just, R. E., and Leathers, H. D. (1994). Land Allocation in HYV Adoption Models: An Investigation of Alternative Explanations. *Am. J. Ag. Econ.* **76**, 535–546.

Vaughan, D., and Chang, T. T. (1992). *In situ* conservation of rice genetic resources. *Econ. Bot.* **46**, 369–383.

Witcombe, J. R. (1999). Does plant breeding lead to a loss of genetic diversity? In *Agrobiodiversity: Characterization, Utilization, and Measurement* (D. Wood and J. Lenné, Eds.). CAB International, Wallingford, UK.

Wood, D., and Lenné, J. (1997). The conservation of agrobiodiversity on-farm: questioning the emerging paradigm. *Biodiversity and Cons.*, **6**, 109–129.

Zimmerer, K. S. (1999). Overlapping patchworks of mountain agriculture in Peru and Bolivia: Toward a regional–global landscape model. *Human Ecol.* **27**, 135–165.

AGRICULTURE, SUSTAINABLE

G. Philip Robertson and Richard R. Harwood
Michigan State University

GLOSSARY

bush-fallow cropping system An agronomic system in which soil fertility is maintained by allowing native vegetation to regrow following several years of cropping.

conservation tillage Tillage that reduces soil disruption in order to conserve organic matter and water and reduce erosion.

ecosystem services Services provided by an ecosystem to the organisms inhabiting the system or living nearby.

legumes Plants belonging to the family Leguminosae, many of whose members can form symbiotic associations with nitrogen-fixing bacteria.

mixed farming system An agronomic system that incorporates different combinations of herbaceous crops, trees, and animals within a single farm unit.

nitrogen fixation The transformation of atmospheric nitrogen into a form usable by plants.

resource conservation The protection and enhancement of resources on which sustainability depends.

rotation The sequence of crops grown in a single field.

soil organic matter Material in soil containing organic carbon derived from the decomposition of mainly plant residue.

trophic complexity The number and types of organisms that feed at different trophic levels within a community; also known as food web complexity.

SUSTAINABLE AGRICULTURE describes a food and fiber production system that is economically viable, environmentally safe, and socially acceptable over long periods.

I. ELEMENTS OF SUSTAINABILITY

Strictly defined, sustainability denotes any system capable of persisting. Because persistence depends on scale, the term "sustainable agriculture" is entirely scale dependent. An agricultural practice that is sustainable at the scale of an individual field may lack sustainability at the larger farm scale if the inputs required to maintain stable production eventually exceed the capacity of the farm to provide them. Likewise, farm-scale sustainability must be evaluated in the context of local and global regions, and global sustainability requires consideration of global output vs the long-term costs of that output and the ability of global resources to accept those costs.

Of course costs are not simply economic but also social and environmental: If a cropping system requires

Encyclopedia of Biodiversity, Volume 1

large inputs of fertilizer that leak from the system to pollute groundwater drinking supplies and distant coastal fisheries, a system that may be sustainable at the field scale becomes unsustainable at the farm and regional scale—even though the long-term supply of fertilizer is stable and the economic cost of fertilizer is easily borne by higher grain production.

Sustainable agriculture must therefore be defined not just in terms of its long-term economic productivity but also in terms of its environmental and social costs. Many of these costs are value-laden. For a society that values family farms, for example, sustainability must be evaluated in terms of the impact of farming practices on the social structure of rural communities: Do the capital expenses associated with large-scale production systems exclude small producers? For a society that values biodiversity, sustainability must be evaluated in terms of the impact of farming practices on wildlife health and habitat. These considerations take sustainability arguments into sometimes contentious territory because different evaluators may have different social or cultural values. What is socially acceptable in one nation or to one segment of society may be socially unacceptable to another, and these differences must be clearly defined when evaluating sustainability.

Therefore, it is possible to define sustainable agriculture as any suite of agronomic practices that are

- economically viable,
- environmentally safe, and
- socially acceptable.

These elements provide an operational definition of agricultural sustainability. There are many ways to blend them into an overarching definition, and many authors—and policymakers—have done so. One should always keep in mind the subjective nature of these amalgamations, however, and that sustainability is a relative term. Although it may be difficult to foresee what ultimately will be sustainable, one can usually judge whether one set of practices is likely to be more sustainable than another at a societal scale.

II. DEVELOPMENT OF THE SUSTAINABILITY CONCEPT

Sustainable agriculture as a descriptive term evolved toward common usage in the United States in the early 1980s as a mixture of concepts, ideas, values, and development direction that many believe is a vision for what agriculture should be. As with most visions, it has had strong impetus from a small group of critics of the conventional paradigm, notably Wes Jackson in his 1980 book *New Roots for Agriculture*, and by 1984 at a conference at Michigan State University the term was in widespread use (Edens et al., 1985). By 1991, the term had been fully "legitimized" as evidenced in the National Research Council's (NRC) documentation of the many U.S. Department of Agriculture, university, and other programs in sustainable agriculture throughout the United States and abroad (NRC, 1991).

Many of the philosophical roots for sustainable agriculture that coalesced in the 1980s were traced from as far back as the Greek and Roman philosophers by Harwood (1990). The public agenda debate over sustainable agriculture of the 1980s embodied many of the older concepts but certainly not any one in its entirety. The notion of land ownership, cultivation, and personal connectedness to the land as a grounding for personal responsibility, morality, and sense of purpose can be found in the writings of English philosopher John Locke in the 1600s, those of Thomas Jefferson in the late 1700s, and recently in the poetry and books of Wendell Berry (1977). Wes Jackson applied the concept to the 1990s in his book *Becoming Native to This Place* (1994), which focused on homecoming and on being native, i.e., developing an ecological literacy in a holistic sense, through being, living, and having personal experience with one's ecosystem. The values in having a sense of place, a connectedness, have changed from an earlier focus on moral and political values; these values have nevertheless become a part of the vision.

The concepts of conservation and natural resource preservation are woven throughout sustainability, derived from the work of Aldo Leopold and others earlier in this century. Louis Bromfield (1947) and the prolific writings of Faulkner (1943, 1952) had major impact. The writings of Faulkner had a significant influence on both J. I. Rodale and his son, Robert, who carried the notion further to one of regenerative agriculture (Rodale, 1983), defined as an agriculture which not only maintains the natural resource base but also restores and increases its productive potential.

Organic, biodynamic, and the many schools of thought and broad literature surrounding these terms made major contributions to the sustainability debate. The "humus farming" school, with its focus on management of soil organic matter and the interconnectedness of soil health, plant health, and that of animals and humans, was a foundation for this philosophy. Several schools of organic practice originated in Europe and England, but the movement was popularized as organic agriculture in the United States by J. I. Rodale in his widely read book *Pay Dirt* (1945).

Each of these philosophical roots contained the notion of a holistic structure of agriculture—an interconnectedness between people, other living organisms, and the soil. Organic agriculture today continues a focus on "the living soil," on optimizing the use of biological processes and on avoiding the use of synthetic chemicals and fertilizers. Most sustainable advocates agree with a biological focus and hope to reduce but not necessarily eliminate chemical use.

Major impetus for the new vision also evolved from the 1960s era of intense agricultural development. The magnitude of the chemical revolution with its newly available pesticides, herbicides, and fungicides plus the availability of concentrated, more easily handled fertilizers very much narrowed the agricultural development focus to the high productivity of major cereal crops. The very real specter of massive starvation in China and hunger and food deficits in India, Bangladesh, and many other countries drove the agricultural development of the 1960s. The low cost of oil and gas and prosperous northern economies all combined for infrastructure development, high agricultural inputs, and the well-known green revolution of this period. A backlash of environmental impact, embodied in Rachel Carson's *Silent Spring* (1962), the oil crisis of the early 1970s, and concern over a growing gap between rich and poor people and nations all set the stage for the new vision. The senescence and decline of rural communities in the United States during the 1970s and 1980s added urgency to the social dimensions of agriculture. The narrowness and focus of the development debate of the 1960s set the stage for a new, broader vision that included revisiting the old and much introspection amid assurances of decreasing real food prices and near-term global food security.

The 1980s and early 1990s produced a plethora of sustainable agriculture development writings. *Future Horizons: Recent Literature in Sustainable Agriculture* (Hegyes and Francis, 1997) is an excellent review of this broad literature. The terms "low-input sustainable" and "alternative agriculture" appeared frequently and are still used by many non-American authors, particularly with reference to organic production systems. The *American Journal of Alternative Agriculture* was founded in the mid-1980s, and in 1989 the NRC's report, *Alternative Agriculture*, brought the name "alternative agriculture" to prominence. Alternative agriculture was defined, in these sources, as any food or fiber production that has

- a more thorough incorporation of natural processes;
- reduced use of off-farm inputs, with less harm to the environment and consumers;
- a more productive use of the biological and genetic potential of plants and animals;
- a better match between cropping patterns and the physical capacity of lands; and
- an improved emphasis on conservation of soil, water, energy, and biological resources.

Alternative agriculture is not synonymous with organic agriculture (which completely avoids synthetic chemical inputs), but they share many of the same farm management practices and approaches.

The 1980s and 1990s have led us from a primary focus on engineering and chemistry in agriculture toward a greater emphasis on biology—from an age of "alchemy" to the age of "algeny" (Rifkin, 1983). The adoption of the term "agroecology" signaled a reemphasized trend in holistic thinking and analysis. A shift in emphasis within the field of ecological science toward managed ecosystems in the late 1980s added significant perspective. Entomologists, agronomists, and other scientists began to use an ecological, process-oriented approach. Stephen Gliessman's *Agroecology: Ecological Processes in Sustainable Agriculture* (1998) is one of the latest books on this topic. The infusion of ecological thinking has added clarity to our understanding of overlays of subsystems, of spatial hierarchies, and of how complexity makes up an agricultural ecosystem. It has added to the understanding of multiple functions of an agricultural system and to the notion of ecosystem services. Most importantly, agroecology has taken the analysis of sustainability to a process level of understanding that allows us to understand gradients of change in production systems across time and space.

Giving greater voice in development direction to farmers has been another significant dimension in sustainable agriculture. It was very apparent that much, if not most, of alternative agriculture had originated with farmers. Philosophical direction had been heavily influenced by farmer–writers. The resurgence of on-farm research in the developing world of the 1970s (Harwood, 1979) was followed by similar emphasis with farmer collaborators in the United States in the 1980s and 1990s. Farmers have been increasingly invited to serve on steering committees and research grant award committees for sustainable agriculture projects.

Farm family well-being and that of their rural communities has been another major area of merger and inclusion in the sustainability vision. The many links and interdependencies between human and community development and sustainable agriculture are being considered in the structure and design of food systems. In 1986, Dahlberg presented a comprehensive and thought-provoking collection of writings on social, economic, and structural issues. The importance of farm size, community interaction, and the global structure of the food system is seen to be critical to both social and economic well-being. Heffernan, as cited in the Dahlberg book and in more recent publications (Heffernan, 1997), forcefully makes the point that with globalization of capital markets and the resulting centralization of control and ownership (of both input supply and product handling and processing) have come a reduction in market competition, a shift in balance of economic power away from the producer, and a replacement of farm-level production instability with greater macro-economic instability in the marketplace. Much of the current literature on the social dimensions of sustainable agriculture is found in *Agriculture and Human Values*, the *Journal of the Agriculture, Food,* and *Human Values Society*.

Economic dimensions of sustainable agriculture have been typically associated with whole-farm studies comparing organic with "conventional" farms in the late 1970s, exemplified by paired comparison studies in the Midwest (Locheretz *et al.*, 1981). Other, single-farm studies have been reported, such as in the case studies of the NRC's report *Alternative Agriculture* (1989). Most of these studies have shown organic farms, in most years, to be as profitable as conventional farms. Most of the sustainable agricultural production research of the past two decades has focused on comparisons of crop rotations and use of cover crops and other systems component practices.

There has been a growing crescendo of voices critical of the failure to account for the side effects, the "external costs" of conventional agriculture. These external factors impact communities, the environment, and human health. There is increasing criticism of the emerging structure, which includes farm scale, the patterns of movement of food, and the corporate concentration of input supply and processing on a global scale. A significant component of the sustainable agriculture debate concerns the desirability of some portion of foods being of local origin, with the size of that portion related to location and community development status (Shuman, 1998; Harwood, 1998).

Long-standing schools of thought and practice in agriculture, influenced by changes in science and technology and in food system structure, have thus provided much of the content of the present-day sustainable agriculture agenda or vision. The amalgam of ideas and component factors has provided an extremely rich background from which development direction can be modified. The breadth of the agenda and the level of dissatisfaction with the current system point to a very major underlying problem of the global, monetarily driven process that is directing and fueling current change. Most of the sustainable agricultural debate concerns differences in goals and in ethical and value dimensions between farmers, agriculture as a sector, and national, international, and global interests as pointed out so clearly by Dahlberg (1985). Many of the resources used and managed by agriculture and many of the services and outputs from agriculture that are critical to ecological and human well-being lie outside the monetary process that is currently driving global agriculture. If the marketplace does not put value on those dimensions and the political process either cannot or will not value them, there is a high level of disarray. Much of today's sustainable agriculture debate revolves around these value differences. Many of these sustainability issues are deeply imbedded in the current public debate over genetically modified organisms.

III. INDICATORS OF SUSTAINABILITY

What agricultural practices are sustainable? This is an area of intensive research. Sustainable practices must meet the three criteria defined in Section I: They must be economically viable, environmentally safe, and socially acceptable. There is no single prescription for sustainability; sustainable practices will vary by cropping system, local environment, and socioeconomic system. Nonetheless, emerging research results suggest that locally sustainable systems tend to be more resource conservative than less sustainable systems and tend to rely less on external subsidies and more on internal ecosystem services.

A. Resource Conservation

Resource conservation means that those resources on which sustainability depends are conserved and even enhanced by agronomic management. Soil organic matter is a good example of an ecosystem resource that is easily reduced without effective management. Soil organic matter declines rapidly in almost all cropping systems following initial cultivation—typically to 40–60% of original values within a few decades. However, soil organic matter is a valuable resource, providing habitat and energy for soil organisms, a soil structure favorable for plant growth and water retention, and a chemical structure favorable for nutrient retention.

The loss of soil organic matter is often associated with a need for greater external inputs. Cropping practices that conserve or enhance soil organic matter buildup will invariably enhance the environmental and often the economic sustainability of cropping systems. Crops grown in high-organic matter soils have a better water and nutrient environment than similar crops grown in soils that are depleted in organic matter, and thus they may require fewer external inputs for the same productivity. Additionally, less soil erosion and lower runoff from high-organic matter soils better protects downstream environments from agronomic impact. Therefore, cropping practices that conserve soil organic matter can be considered more sustainable than those that do not.

Often, however, there are trade-offs that require any specific conservation effort to be evaluated in the overall context of sustainability. For example, conservation tillage typically slows or stops soil organic matter loss and thus can be considered a resource-conserving, sustainable cropping practice. However, tillage controls weeds in cropping systems, and in the absence of tillage weed control is typically left to herbicides, which have environmental and economic costs different from those of tillage. Is the maintenance of soil organic matter as sustainable in light of a more intensive reliance on herbicides? Ideally, such trade-offs can be minimized. For example, winter cover cropping can also reduce soil organic matter loss and additionally can reduce nitrate leaching and suppress weeds, without the need for additional herbicide. Nevertheless, each cropping practice must be evaluated in a whole-system context to adequately evaluate its contribution to a system's sustainability.

B. Ecosystem Services

Ecosystem services are those services provided by an ecosystem to the organisms inhabiting the system or living nearby. Unmanaged systems provide such services as a matter of course. Farms can likewise provide such services to organisms within the farming system as well as to organisms in the surrounding landscape and to local communities. Services such as pollination, water retention and groundwater recharge, a particular light environment, or food sources—whether provided directly by other organisms or indirectly by their effects on local environmental conditions—are integral to the functioning of healthy ecosystems. In modern cropping systems many services provided by the original ecosystem prior to its conversion to agriculture have been suppressed or ignored in favor of services provided by external inputs. In a nitrogen-poor native ecosystem, biological nitrogen fixation by native legumes such as clover (*Trifolium* spp.) might be a principal source of fixed nitrogen; modern cropping systems rely almost exclusively on industrially fixed nitrogen provided as inorganic fertilizer. In a native or unmanaged system, insect herbivory is suppressed largely by trophic and structural complexity that enables insect and vertebrate predators to keep plant pests at bay. In most modern systems insect pests are controlled with insecticides, which also kill insect predators. Managing a cropping system with legumes or with greater plant diversity (either within fields or adjacent to field edges) would allow the ecosystem to provide more of the services now provided via external inputs. Legume cover crops can reduce the need for external nitrogen, and greater plant diversity can provide the structural complexity and refugia needed to support predator populations in otherwise monospecific landscapes.

Just as for practices intended to enhance resource conservation, practices established to reintroduce or enhance existing ecosystem services need to be evaluated on the basis of their total net contribution to sustainability. Although nitrogen fixation by legumes can lower the need for fertilizer inputs and benefit soil organic matter buildup as well as provide winter habitat for predaceous insects, there is no evidence that legume-fixed nitrogen is conserved more tightly than fertilizer-derived nitrogen. Thus, there may be no downstream environmental benefit associated with this ecosystem service. Likewise, animal manure produced on-farm and recycled back to the field may be less conserved than fertilizer nitrogen if the manure is added out of synch with plant nutrient demands. Ongoing research is helping to identify ways in which management can add ecosystem services that both enhance resource availability and reduce the environmental costs of agriculture. At the societal scale there is ongoing debate on how to value services provided by farms to their neighboring communities. Should farmers be compensated for managing their land in ways that provide services to local, regional, and national communities? Such payments occur in some parts of Canada, the United States, and Europe today.

IV. EXAMPLES OF SUSTAINABLE CROPPING SYSTEMS

A. Bush Fallow Rotation Systems

Perhaps the best documented example of a locally sustainable cropping system is the bush-fallow rotation, also known as swidden and slash and burn agriculture, indigenous to many cultures prior to the advent of continuous cropping systems several hundred years ago, and still evident in the humid tropics today. In the absence of population change, the bush-fallow system allows a tract of forest or savanna to provide food with few subsidies other than human labor.

In these systems, a small section of native vegetation is cut and cropped. Crop nutrient needs are met by the decomposition of soil organic matter and perhaps by leguminous crops. Sufficient pest protection is provided by crop rotation, complex crop mixtures, and the proximity of fields to native vegetation. Weed suppression is performed by hand.

Once soil nutrients are depleted to levels that significantly compromise crop productivity, the plot is abandoned to "bush fallow" and another plot is cleared from native vegetation and cropped. Meanwhile, the newly fallowed plot is undergoing secondary succession with attendant recovery of soil nutrient stores. By the time several more plots have been sequentially cut, cropped, and fallowed, the original plot will have recovered much of its original fertility and be ready to be cleared and cropped again.

Such agronomic systems are sustainable indefinitely as long as each cropped area is allowed sufficient time to recover its original fertility. However, when land becomes scarce because of development or population growth or both, the system can quickly fail. Native vegetation brought out of fallow too quickly will provide soil fertility for only a portion of the former cropping periods, so the crop portion of the rotation will either be shorter or yield less, forcing more of the native vegetation to be brought out of fallow earlier than planned in order to feed a growing population. Eventually, little native fallow will remain and crops will be grown continuously on soils that now lack much of their former fertility. One of today's greatest agronomic challenges is providing nitrogen and phosphorus to cropping systems that until recently have been in bush-fallow rotations, especially in sub-Saharan Africa where fertilizer is largely unavailable and most food is grown on small holdings of a few hectares. The maintenance of soil quality and adequate levels of soil organic matter to provide it are major concerns of tropical agronomists.

B. Mixed Farming Systems

The successor to simple bush-fallow systems is mixed farming systems that have several production enterprises of different herbaceous crops, trees, animals, or combinations of crops and animals. In less developed or unstable economies requiring a high level of local, community, and farm family self-reliance, the production of a wide array of goods was primarily to meet family and local market needs for an ensured, year-round supply of food, fuel, and building materials. Farm and landscape-level diversity optimized stability within local environments and increased the resiliency of the system to a wide variety of disturbances. The diversity of land use provided a wide range of ecosystem services, including precipitation management, groundwater recharge, wildlife habitat, an environment usually conducive to adequate pest–predator balance, and some mitigation of harsh climatic conditions. The mixed plant community provided shade, wind protection, privacy, and many other, often seasonal, assorted products and services. This range of outputs has recently been termed the multifunctional character of agricultural land (Food and Agricultural Organization, 1999) shown in Fig. 1

As infrastructure and markets develop, the need for a broad range of products and services decreases. When the costs of adverse environmental impacts such as groundwater contamination are not internalized, or when farmers are not rewarded for ecosystem services that their farms provide to the community or region, they do not include such values in their farm enterprise unless they are motivated and willing to make an altruistic contribution. Many farmers, in fact, do this now, but ultimately the more narrowly focused economic marketplace rules. Today's farms in highly developed economies frequently have a level of product and land use specialization that is well below an acceptable standard for long-term environmental and resource sustainability (Fig. 1B). In other words, the production base, the environment, and its ecosystem have not been stabilized and are being degraded. With continued market evolution, farmers may be increasingly compensated for the full range of ecosystem services as well as actual product output that they provide (Shuman, 1998; Soule and Piper, 1992) (Fig. 1C).

A more immediate incentive is to add crops and/

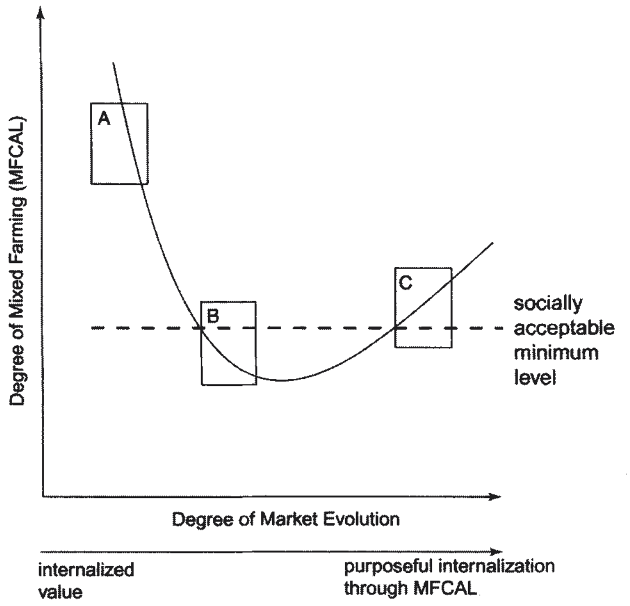


FIGURE 1 Changes in the purposeful development of the multifunctional character of agricultural land (MFCAL) as markets evolve (adapted from L. Fresco, personal communication).

or livestock to provide integration efficiencies. These efficiencies include the increase in yield of one crop following another, the savings in nutrient inputs, or the reduction in pest control costs (Table I).

TABLE I

Michigan Maize, Soybean, and Wheat Rotation Efficiencies

System, advantages
Maize after beans
30 kg/ha nitrogen credit
No rootworm scouting or control costs
6–10% yield advantage
Maize after soybeans dry beans wheat (Michigan, 2nd, 3rd year of rotation)
No nitrogen credit (since maize follows wheat)
No rootworm control costs
Window for perennial weed control (either mechanical or chemical)
Greater than 10% yield advantage (because of the preceding bean/wheat sequence)
Maize after wheat plus frost-seeded clover
40 kg/ha nitrogen credit (60–70 kg N/ha with pre-sidedress nitrogen test)
No rootworm control costs
At least 15% yield advantage
30–50% yield advantage if the farm is organic, where maize-after-maize is not advisable

An increasing amount of information on the efficiencies of specific technologies for integration is becoming available in the scientific literature. The reduction in input requirements is often a key part. There is less direct research information on the relationship of many of these practices on environmental loading. An exception is the wealth of data on reduced soil erosion as a result of reduced or zero tillage. Currently, the predictive models of loss of pesticides, nutrients, or crop or animal residues are rudimentary. Direct measurements of loss from alternative rotations and use of cover crops are very difficult, expensive, and location specific. These rotation and cover crop practices are widely acknowledged as being fundamental to sustainability. Their efficiencies are being quantified with respect to yield, input reduction, and soil quality and the prevention of soil loss. Michigan data show, for instance, that wheat in rotation loses less than 20 kg N/ha per year via groundwater leaching. Well-fertilized continuous corn averages 50 kg N/ha per year. Most U.S. farmers use at least a two-crop rotation.

Animal integration in crop systems is declining in the United States. Poultry and turkeys are increasingly produced in specialized production facilities not located on the farms where their feed is produced. They are usually located in areas where agricultural land is available for manure application, often on a contract "disposal" basis. The level of crop or animal diversity that is appropriate on a farm to balance the market forces for specialization with the need for biological efficiency and ecosystem maintenance is very situation specific. As enterprise integration increases with an effective level of appropriate enterprises and their effective management (Fig. 2, technology T_2), agricultural output can be maintained at a much higher level for a given amount of ecosystem disturbance. In other words, sustainable agriculture can maintain productivity at a much lower level of ecosystem disturbance. Very large-scale operations tend to have less diversity, in part because of the greater difficulty of managing diverse enterprises. Crop and animal management requires numerous and often frequent decisions to be made as conditions change that are often stimulated by visual, difficult to measure changes. The frequent presence and sensitivity of the manager, the experience in production management, and the ability to make decisions place limits on the scale of highly diversified operations. Every farm owner experiences this tension.

On a global scale, under conditions of high population density, small farms, and the need for producing a wide array of products in often marginal production environments, a very diverse type of farm enterprise

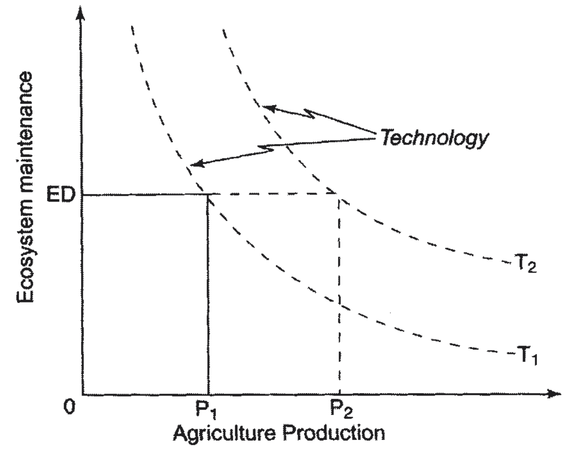


FIGURE 2 Effect of technology selection on the relationship between agricultural production and ecosystem maintenance. ED, environmental disturbances; T_1, inappropriate technology; T_2, more appropriate technology; P_1, agricultural production with technology T_1; P_2, agricultural production with technology T_2 (adapted from NRC, 1993).

mix is common. Trees become a very important part of farm productivity in the higher rainfall areas where they are a part of the native vegetation. Animals are more often than not a part of the enterprise because they consume crop residues and add significantly to overall productivity (Fig. 3). In most developing countries there is very little area of undisturbed forest, and in only a few in which large tracts of land remain is there cattle only on farms. In the humid tropics the mixture of crops, trees, and animals (agrisilvo-pastoral systems) represent the great majority of farms. Where human population is relatively high (>300–500 persons per square kilometer) in rural areas, if there is poverty combined with modest levels of rainfall (less than 1000 mm per year) and/or cool temperatures for part of the year, fuel for cooking and heating becomes a problem. Resource degradation and loss of production potential often occur as the standing stocks of carbon (particularly in trees) and eventually the soil carbon stocks are reduced as crop and animal residues are burned. The system rapidly loses crop nutrient holding and recycling capacity, and its ability to intercept and retain rainfall decreases.

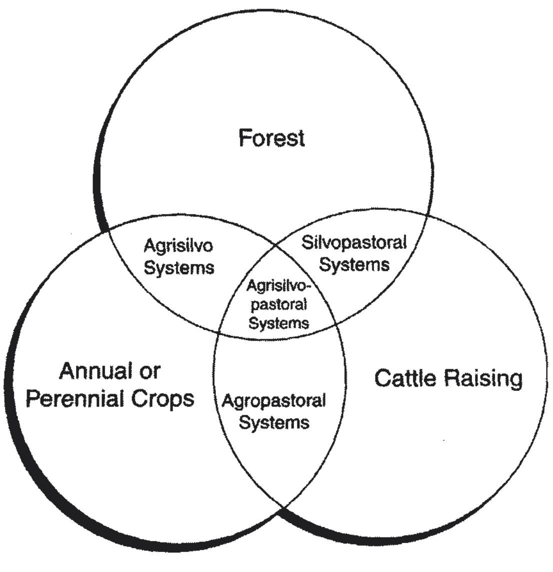


FIGURE 3 Combinations of enterprises in mixed farming systems (redrawn from Serrão and Homma, 1993).

The immigrant Japanese farms of the Brazilian area are an interesting example of high diversity, maintenance of a high-standing stock of carbon, and extremely high market value on a small land area (Fig. 4). This is similar in many ways to Indonesian agroforestry systems in Central Sumatra. Most developing country mixed farms have a larger portion of cash grain crops.

V. CONCLUSIONS

Our global food and fiber production systems are undergoing an enormous transformation, driven by rapid advances in the sciences of biotechnology, engineering, and food processing and chemistry. The increasing centralization in the manufacture of agricultural inputs and in the collection and processing of food is causing huge economic and social change. Many social and political values are being challenged. The global marketplace forces for product uniformity and the geographical concentration of its production are driving farmers toward a level of specialization that results in their farms having much less diversity of crop and animal enterprises on the landscape than that desirable for the maintenance of many ecosystem services. Markets have not yet matured to adequately value these services, especially those that affect environmental quality, nor in most places have governments established disincentives in order to protect them. The greatest challenge to sustainability, with its many economic, environmental, and social dimensions, is the lack of public awareness, vision, and will to implement necessary changes. In some cases, research is needed to clarify the value of alternative strategies and to provide additional options for sustainable management.

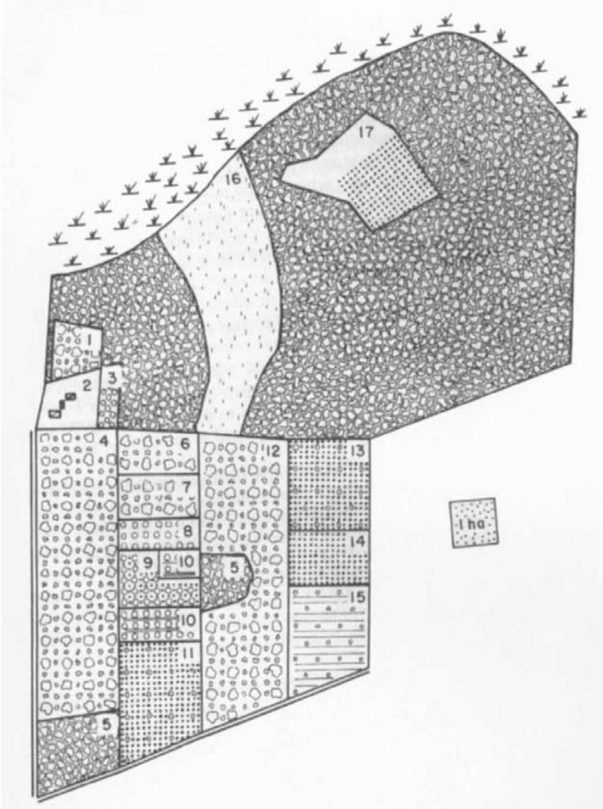


FIGURE 4 Land use on a representative Nippo-Brazilian farm. 1, Cacao, erythrina; 2, household area; 3, coconut, citrus, mangosteen, and graviola; 4, cacao, erythrina, andiroba, and Brazil nut; 5, secondary forest regeneration; 6, cacao, vanilla, palheteira, and freijó; 7, cacao and paricá; 8, rubber trees; 9, rubber trees, black pepper, and cacao; 10, rubber trees and passion fruit; 11, black pepper and cacao; 12, cacao, banana, and *Cecropia* sp.; 13, black pepper and cupuaçu; 14, black pepper; 15, passion fruit and cupuaçu; 16, pasture grasses; 17, black pepper and clearing (reproduced with permission from Subler and Uhl, 1990).

For the present, and in the aggregate, the world has enough food. In many if not most cases it is not being produced sustainably, i.e., in a manner that is economically viable, environmentally benign, and socially acceptable to many who are affected by its production. On the other hand, new research is showing that sustainable cropping systems can be designed to operate effectively, using ecological knowledge to substitute for some of the management options now provided by external inputs, and in a way that has a less adverse environmental and social impact than conventional management.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • AGRICULTURE, TRADITIONAL • BIODIVERSITY AS A COMMODITY • ECOSYSTEM SERVICES, CONCEPT OF • HERBICIDES • NITROGEN AND NITROGEN CYCLE • PESTICIDES, USE AND EFFECTS OF • SOIL CONSERVATION

Bibliography

Altieri, M. (1995). *Agroecology: The Science of Sustainable Agriculture.* Westview, Boulder, CO.

Berry, W. (1977). *The Unsettling of America.* Sierra Club, San Francisco. (Reprinted in 1986)

Bromfield, L. (1947). *Malabar Farm.* Balantine, New York.

Carroll, C. R., Vandermeer, J. H., and Rosset, P. M. (1990). *Agroecology.* McGraw-Hill, New York.

Carson, R. (1962). *Silent Spring.* Houghton Mifflin, Boston.

Dahlberg, K. A. (1985). Ethics, values and goals in agricultural systems and agricultural research. In *Sustainable Agriculture and Integrated Farming Systems: 1984 Conference Proceedings* (T. C. Edens, C. Fridgen, and S. L. Battenfield, Eds.), pp. 202–218. Michigan State Univ. Press, East Lansing.

Dahlberg, K. A. (Ed.) (1986). *New Directions for Agriculture and Agricultural Research: Neglected Dimensions and Emerging Alternatives.* Rowman & Allanheld, Totowa, NJ.

Edens, T. C., Fridgen, C., and Battenfield, S. L. (Eds.) (1985). *Sustainable Agriculture and Integrated Farming Systems: 1984 Conference Proceedings.* Michigan State Univ. Press. East Lansing.

Faulkner, E. (1943). *Plowman's Folly and a Second Look.* Island, Washington, DC.

Faulkner, E. (1952). *Soil Development.* Univ. of Oklahoma Press, Norman.

Food and Agricultural Organization (FAO) (1999). *Conference on the Multifunctional Character of Agricultural Land.* Maastricht, The Netherlands. September 12–17. FAO, Rome.

Francis, C. A., Flora, C. B., and King, L. D. (Eds.) (1990). *Sustainable Agriculture in Temperate Zones.* Wiley, New York.

Gliessman, S. R. (1998). *Agroecology: Ecological Processes in Sustainable Agriculture.* Ann Arbor Press, Chelsea, MI.

Harwood, R. R. (1979). *Small Farm Development: Understanding and Improving Farming Systems in the Humid Tropics.* Westview, Boulder, CO.

Harwood, R. R. (1990). A history of sustainable agriculture. In *Sustainable Agricultural Systems* (C. A. Edwards, R. Lal, R. Madden, R. Miller, and G. House, Eds.), pp. 3–19. Soil and Water Conservation Society, Ankeny, IA.

Harwood, R. (1998). Directions for farm change: Bringing it all together. In *Michigan Field Crop Ecology: Managing Biological Processes for Productivity and Environmental Quality* (M. A. Cavigelli, S. R. Deming, L. K. Probyn, and R. R. Harwood, Eds.), Extension Bulletin No. E2646, pp 80–86. Michigan State Univ. Press, East Lansing.

Heffernan, W. D. (1997). Domination of world agriculture by transnational corporations. In *For All Generations: Making World Agriculture More Sustainable* (J. P. Madden and S. G. Chaplow, Eds.), pp. 173–181. World Sustainable Agriculture Association, Glendale, CA.

Hegyes, G., and Francis, C. A. (Eds.) (1997). *Future Horizons: Recent Literature in Sustainable Agriculture.* Center for Sustainable Agricultural Systems, University of Nebraska, Lincoln.

Jackson, W. (1980). *New Roots for Agriculture.* Univ. of Nebraska Press, Lincoln.

Jackson, W. (1994). *Becoming Native to This Place.* Univ. Press of Kentucky, Lexington.

Lockeretz, W., Shearer, G., and Kohl, D. H. (1981). Organic farming in the Corn Belt. *Science* **211**, 540.

National Research Council (1989). *Alternative Agriculture.* National Academy Press, Washington, DC.

National Research Council (1991). *Sustainable Agriculture Research and Education in the Field: A Proceedings.* National Academy Press, Washington, DC.

Rifkin, J. (1983). *Algeny.* Viking, New York.

Rodale, J. I. (1945). *Pay Dirt.* Rodale Press, Emmaus, PA.

Rodale, R. (1983). Breaking new ground: The search for sustainable agriculture. *The Futurist* **1**, 15.

Serrão, E. A. S., and Homma, A. K. O. (1993). Brazil. In *National Research Council. Sustainable Agriculture and the Environment in the Humid Tropics*, pp. 265–351. National Academy Press, Washington, DC.

Shuman, M. H. (1998). *Going Local: Creating Self-Reliant Communities in a Global Age.* Free Press, New York.

Soule, J. D., and Piper, J. K. (1992). *Farming in Nature's Image: An Ecological Approach to Agriculture.* Island, Washington, DC.

Subler, S. J., and Uhl, C. (1990). Japanese agroforestry in Amazonia: A case study in Tome-Acu, Brazil. In *Alternatives to Deforestation: Steps Toward Sustainable Use of the Amazon Rain Forest* (A. B. Anderson, Ed.), pp. 152–166. Columbia University Press, New York.

AGRICULTURE, TRADITIONAL

Miguel A. Altieri
University of California, Berkeley

GLOSSARY

agroecosystem A simplified natural ecosystem subjected to exploitation for purposes of food and fiber production.

biodiversity Diversity of microbial, animal, and plant species in an ecosystem that performs distinct ecological functions and services.

ethnoecology Study of the various forms of traditional environmental knowledge that are characteristic of specific ethnic groups and that translate into natural resource management.

polyculture Intensive growing of two or more crops either simultaneously or in sequence on the same piece of land.

sustainable agriculture Form of agriculture that is environmentally sound, culturally sensitive, socially acceptable, and economically viable.

traditional agriculture Indigenous form of ecologically based agriculture resulting from the coevolution of local cultural and environmental systems.

TRADITIONAL AGRICULTURE IS AN INDIGENOUS FORM OF FARMING that results from the coevolution of local cultural and environmental systems. It exhibits a high level of ecological rationale expressed through the intensive use of local knowledge and natural resources, including the management of agrobiodiversity in the form of diversified agricultural systems.

I. INTRODUCTION

One of the salient features of traditional farming systems throughout the developing world is their high degree of biodiversity. These traditional systems have emerged over centuries of cultural and biological evolution and represent the accumulated experiences of indigenous farmers interacting with the environment without access to external inputs, capital, or modern scientific knowledge (Chang, 1977; Grigg, 1974). Using inventive self-reliance, experiential knowledge, and locally available resources, traditional farmers have often developed farming systems that generate sustained yields (Har-

Encyclopedia of Biodiversity, Volume 1

wood, 1979). In Latin America alone, more than two and a half million hectares are under traditional agriculture in the form of raised fields, polycultures, and agroforestry systems, documenting the successful adaptation of these farming practices to difficult environments (Altieri, 1991).

Many of these traditional agroecosystems, still found throughout the Andes, Meso-America, and the lowland tropics, constitute major *in situ* repositories of both crop and wild plant germplasm. From an agroecological perspective, these agroecosystems can be seen as a continuum of integrated farming units and natural or seminatural ecosystems where plant gathering and crop production are actively pursued. Plant resources are directly dependent on management by human groups; thus, both species and genetic diversity have evolved in part under the influence of farming practices shaped by particular cultures and the forms of sophisticated knowledge they represent (Nabhan, 1983).

Perhaps the key to understanding how traditional farmers maintain, preserve, and manage biodiversity is to recognize the complexity of their production systems. Today, it is widely accepted that indigenous knowledge is a powerful resource in its own right and complementary to knowledge available from Western scientific sources. Therefore, in studying such systems, it is not possible to separate the study of agricultural biodiversity from the study of the culture that nurtures it.

This article explains the features of the biodiversity inherent in traditional agroecosystems, and the ways in which farmers apply local knowledge to manage such biodiversity to satisfy subsistence needs and to obtain ecological services. Traditional agriculture is rapidly disappearing in the face of major social, political, and economic changes. The conservation and management of this agrobiodiversity will be possible only if they are linked to the preservation of the cultural diversity and economic viability of the local farming populations.

II. BIODIVERSITY FEATURES OF TRADITIONAL AGROECOSYSTEMS

Traditional farming systems commonly support a high degree of plant diversity in the form of polycultures and/or agroforestry patterns (Chang, 1977; Clawson, 1985). This strategy of minimizing risk by planting several species and varieties of crops stabilizes yields over the long term, promotes diet diversity, and maximizes returns even with low levels of technology and limited resources (Harwood, 1979). Such biodiverse farms are endowed with nutrient-enriching plants, insect predators, pollinators, nitrogen-fixing and nitrogen-decomposing bacteria, and a variety of other organisms that perform various beneficial ecological functions.

Traditional multiple-cropping systems provide as much as 15–20 percent of the world food supply (Francis, 1986). Polycultures constitute at least 80 percent of the cultivated area in West Africa and predominate in other parts of Africa as well (Norman, 1979). At the same time, much of the production of staple crops in the Latin American tropics occurs in polycultures. More than 40 percent of the cassava, 60 percent of the maize, and 80 percent of the beans in the region grow in mixtures with each other or other crops (Francis, 1986). Polycultures are also very common in parts of Asia where upland rice, sorghum, millet, maize, and irrigated wheat are the staple crops. Lowland (flooded) rice is generally grown as a monoculture, but in some areas of Southeast Asia farmers build raised beds to produce dryland crops amid strips of rice (Beets, 1982).

Tropical agroecosystems composed of agricultural and fallow fields, complex home gardens, and agroforestry plots commonly contain well over 100 plant species per field, and these are used as construction materials, firewood, tools, medicines, livestock feed, and human food. Examples include multiple-use agroforestry systems managed by the Huastecs and Lacondones in Mexico, the Bora and Kayapó Indians in the Amazon River basin, and many other ethnic groups who incorporate trees into their production systems (Wilken, 1977).

In the Latin American tropics, home gardens are a highly efficient form of land use, incorporating a variety of crops with different growth habits. The result is a structure similar to that of tropical forests, with diverse species and a layered physical configuration (Denevan *et al.*, 1984). In Mexico, for example, Huastec Indians manage a number of fields, gardens, and forest plots that may harbor a total of about 300 species. Small areas around their houses commonly average 80–125 useful plant species, mostly native and medicinal plants. Huastec management of the noncrop vegetation in these complex farm systems has influenced the evolution of individual plants and the distribution and composition of the crop and noncrop communities.

In these "forestlike" agricultural systems, nutrient cycles are tight and closed. In traditional shaded coffee plantations (where *Inga* and *Erythrina* are common tree species), the total nitrogen input from the decomposition of shade tree leaves, as well as from litter and symbiotic fixation, can be well over ten times higher than the net nitrogen output in the coffee harvest, which usually averages 20 kg/ha/yr. Clearly, the system amply compensates for the nitrogen loss by harvest with a subsidy from the shade trees. In Mexico, farmers encourage the growth of native leguminous trees in cultivated fields (Wilken, 1977). From Puebla and Tehuacán south through Oaxaca, farms with light to moderately dense stands of mesquite (*Prosopis* spp.), guaje (*Leucaena esculenta*), and guamuchil (*Pithecellobium* spp.) are a familiar sight. Stand density varies from fields with only a few trees to virtual forests with crops planted beneath them. A slightly different practice is found near Ostuncalco, Guatemala, where rigorously pruned sauco (*Sambucus mexicana*) stumps dot maize and potato fields. Sauco leaves and small branches are removed annually, scattered around individual crop plants, and then chopped and interred with broad hoes. Local farmers claim that crop quality and yields in the sandy volcanic soils of this region depend on the annual application of this method (Wilken, 1977).

Many traditional agroecosystems are located in centers of crop diversity, and thus contain populations of variable and adapted landraces as well as wild and weedy relatives of crops (Harlan, 1976). Clawson (1985) described several systems in which tropical farmers plant multiple varieties of each crop; this practice supports both intraspecific and interspecific diversity, and also enhances harvest security. For example, in the Andes, farmers cultivate as many as 50 potato varieties in their fields (Brush, 1982). Similarly, in Thailand and Indonesia, farmers maintain a diversity of rice varieties adapted to a wide range of environmental conditions, and they regularly exchange seeds with each other (Grigg, 1974). The resulting genetic diversity heightens resistance to diseases that attack particular strains of the crop and enables farmers to exploit different microclimates and to derive multiple nutritional and other uses from the genetic variation among the species.

Many plants within or around traditional cropping systems are wild or weedy relatives of crop plants. In fact, many farmers “sponsor” certain weeds in or around their fields that may have positive effects on soil and crops, or that serve as food, medicines, ceremonial items, teas, soil improvers, or pest repellents. In the Mexican Sierras, the Tarahumara Indians depend on edible weed seedlings or “quelites” (e.g., *Amaranthus, Chenopodium, Brassica*) in the early season from April through July, a critical period before crops mature from August through October. Weeds also serve as alternative food supplies in seasons when maize or other crops are destroyed by frequent hail storms (Bye, 81). In barley fields, it is common for Tlaxcalan farmers to maintain *Solanum mozinianum* at levels up to 4500 plants/ha; this yields about 1300 kg of fruit, a significant contribution to agricultural subsistence (Altieri and Trujillo, 1987).

Farmers also derive other benefits from weeds, such as increased gene flow between crops and their relatives. In Mexico, when the wind pollinates maize, natural crosses occur with wild teosinte growing in the field borders, resulting in hybrid plants. Certain weeds are used directly to enhance the biological control of insect pests, as many flowering weeds attract predators and parasites of pests to their pollen and nectar. Other farmers allow weeds such as goosegrass (*Eleusine indica*) in bean fields to repel *Empoasca* leafhoppers, or wild *Lupinus* as a trap plant for the pestiferous scarab beetle (*Macrodactylus* sp.), which otherwise would attack corn (Altieri, 1993).

However, diversity is maintained not only within a cultivated area. Many farmers maintain natural vegetation adjacent to their fields, and thus obtain a significant portion of their subsistence requirements through gathering, fishing, and hunting in habitats that surround their agricultural plots. For the P’urhepecha Indians who live around Lake Pátzcuaro in Mexico, gathering is part of a complex subsistence pattern that is based on multiple uses of their natural resources. These people use at least 224 species of native and naturalized vascular plants for dietary, medicinal, household, and fuel needs (Caballero and Mapes, 1985).

Depending on the level of biodiversity of closely adjacent ecosystems, farmers accrue a variety of ecological services from surrounding natural vegetation. For example, in western Guatemala, the indigenous flora of the higher-elevation forests provide valuable native plants that serve as a source of organic matter to fertilize marginal soils, for each year farmers collect leaf litter from nearby forests and spread it over intensively cropped vegetable plots to improve tilth and water retention. Some farmers may apply as much as 40 metric tons of litter per hectare each year; rough calculations indicate that a hectare of cropped land requires the litter production of 10 ha of regularly harvested forest (Wilken, 1977).

Clearly, traditional agricultural production commonly encompasses the multiple uses of both natural and artificial ecosystems, where crop production plots and adjacent habitats are often integrated into a single agroecosystem.

III. THE COMPLEX NATURE OF TRADITIONAL FARMERS' KNOWLEDGE

Ethnoecology is the study of the natural world knowledge systems of indigenous ethnic people. This knowledge has many dimensions, including linguistics, botany, zoology, craft skills, and agriculture, and is derived from the direct interaction between humans and their environment. In such a system, cognition and perception select the most adaptive and useful environmental information, and this "successful" knowledge is preserved from generation to generation through oral or experimental means. Indigenous peoples' knowledge about soils, climates, vegetation, animals, and ecosystems usually results in multidimensional productive strategies (i.e., the use of multiple ecosystems with multiple species), and these strategies generate (within certain ecological and technical limits) the food self-sufficiency of farmers in a region (Netting, 1993).

Captivated by the ecological intricacies of these traditional agricultural systems, many scientists are now beginning to show interest in them. As scientists search for ways to remedy the deficiencies of modern agriculture, they recognize that indigenous farmers' knowledge may hold vital information for the future of world agriculture. After centuries of cultural and biological evolution, these farmers have developed locally adapted, complex farming systems that have helped them to sustainably manage a variety of environments and to meet their subsistence needs, without depending on modern agricultural technologies.

For many agricultural scientists, four aspects of these traditional knowledge systems are relevant: knowledge of the environment, folk taxonomies, knowledge of farming practices, and the experimental nature of traditional knowledge (Altieri, 1987).

A. Knowledge of the Environment

Indigenous knowledge about the physical environment is often very detailed. Many farmers have developed traditional calendars to control the scheduling of agricultural activities, and many sow according to the phase of the moon, believing that there are lunar phases of rainfall. They also cope with climatic seasonality by utilizing weather indicators based on the phenologies of local vegetation.

Soil types, degrees of soil fertility, and land-use categories are also discriminated in detail. Soil types are commonly distinguished by color, texture, and sometimes taste. Shifting cultivators usually classify their soils based on vegetation cover. In general, peasants identify soil types based on the nature of the peasant's relationship to the land (Williams and Ortiz-Solario, 1981). Aztec soil classification systems were very complex, recognizing more than two dozen soil types identified by origin, color, texture, smell, consistency, and organic content. These soils were also ranked according to agricultural potential, which was used in both land-value evaluations and rural census. Today, Andean peasants in Coporaque, Peru, recognize four main soil classes, where each class has specific characteristics matching the most adequate cropping system (Brush, 1982).

B. Biological Folk Taxonomies

Many complex knowledge systems that are used by indigenous people to group together plants and animals have been well documented (Berlin *et al.,* 1973). The traditional name of a plant or animal usually reveals that organism's taxonomic status, and researchers have found that, in general, there is a good correlation between folk taxa and scientific taxa.

The classification of animals, especially insects and birds, is widespread among indigenous farmers. Insects and related arthropods have major roles as crop pests, as causes of disease, as food, and as medicinal products, in addition to their importance in local myth and folklore. In many regions of the world, agricultural pests are tolerated because they also constitute agricultural products; that is, traditional agriculturalists may consume plants and animals that would otherwise be considered pests (Brokensha *et al.,* 1980).

Ethnobotanies are the most commonly documented folk taxonomies (Alcorn, 1984). The ethnobotanical knowledge of certain campesinos in Mexico is so elaborate that the Tzeltal, P'urepecha, and Yucatan Mayans can recognize more than 1200, 900, and 500 plant species, respectively (Toledo *et al.,* 1985). Similarly, !ko bushwomen in Botswana were able to identify 206 out of 211 plants collected by researchers (Chambers, 1983), while Hanunoo swidden cultivators in the Philippines could distinguish over 1600 plant species (Grigg, 1974).

C. Knowledge of Farming Practices

As more scientific research is conducted, many of the traditional farming practices once regarded as primitive or misguided are being recognized as sophisticated and appropriate. For example, when confronted with specific problems of slope, flooding, drought, pests, diseases, or low soil fertility, small farmers throughout the world have developed unique management systems aimed at overcoming these constraints (Klee, 1980). In general, traditional agriculturalists have adjusted to environmental constraints by concentrating on a few characteristics and processes that incorporate the following structural and functional elements (Gliessman, 1998; Altieri and Anderson, 1986):

a. They combine high species numbers and structural diversity in time and space (through both vertical and horizontal organization of crops).
b. They exploit the full range of microenvironments (which differ in soil, water, temperature, altitude, slope, fertility, etc.) within a field or region.
c. They maintain closed cycles of materials and wastes through effective recycling practices.
d. They rely on the complexity of biological interdependencies, resulting in some degree of biological pest suppression.
e. They rely on local resources plus human and animal energy, thereby using low levels of technology input.
f. They rely on local varieties of crops and incorporate the use of wild plants and animals. Production is usually for local consumption. The level of income is low; thus the influence of noneconomic factors on decision making is substantial.

D. The Experimental Nature of Traditional Knowledge

The strength of traditional people's knowledge is that it is based not only on acute observation but also on trial and error and experimental learning. The experimental approach is very apparent in the selection of seed varieties for specific environments, but it is also implicit in the testing of new cultivation methods to overcome particular biological or socioeconomic constraints. In fact, Chambers (1983) argued that farmers often achieve a richness of observation and a fineness of discrimination that would be accessible to Western scientists only through long and detailed measurement and computation.

Yet only recently has some of this traditional knowledge been described and documented by researchers. The evidence suggests that the finest discrimination develops in communities where the environments have great physical and biological diversity and/or in communities living near the margins of survival (Chambers, 1983). Also, older community members possess greater, more detailed knowledge than younger members (Klee, 1980).

IV. THE ECOLOGICAL SERVICES OF BIODIVERSITY IN TRADITIONAL AGROECOSYSTEMS

In traditional agroecosystems, complex and diversified cropping systems are vital because the interactions among crops, animals, insects, and trees result in beneficial synergisms that optimize soil fertility, pest control, and productivity (Altieri, 1995; Harwood, 1979; Richards, 1985). Among the ecological services are the following.

1. By interplanting, farmers take advantage of the capacity of cropping systems to reuse their own stored nutrients. The tendency of some crops to deplete the soil is counteracted by interplanting other crops that enrich the soil with organic matter. Soil nitrogen, for example, can be increased by incorporating legumes in the crop mixture, and phosphorus assimilation can be enhanced by growing crops with mycorrhizal associations.

2. The complex structure of traditional agroecosystems minimizes crop loss to insect pests through a variety of biological mechanisms. The intercropping of diverse plant species provides habitats for the natural enemies of insect pests as well as alternative host plants for pests. For example, a crop may be planted as a diversionary host to protect other more susceptible or more economically valuable crops from serious damage. The diversity of crops grown simultaneously in polycultures helps prevent the buildup of pests on the comparatively isolated plants of each species. Where shifting cultivation is practiced, the clearing of small plots from secondary forest vegetation also permits the easy migration of natural pest predators from the surrounding forest.

3. Increasing the species and/or genetic diversity of cropping systems is a key strategy to minimize losses from plant diseases and nematodes (types of roundworms that are among the most widespread and damaging of agricultural pests). The mixing of different crop species or varieties can delay the onset of diseases, reduce the spread of disease-carrying spores, and modify environmental conditions such as humidity, light, temperature, and air movement so that they are less favorable to the spread of certain diseases.

4. Many intercropping systems prevent competition from weeds, chiefly because the large leaf areas of their complex canopies prevent sufficient sunlight from reaching sensitive weed species. In general, the extent to which weeds present a problem depends on the type of crops and the proportion of the different species grown, their density, where they are planted, the fertility of the soil, and management practices. Weed suppression can be enhanced in intercrop systems by adding crop species that inhibit weed germination or growth. Crops such as rye, barley, wheat, tobacco, and oats release toxic substances into the environment, either through their roots or from decaying plant material. Such toxins inhibit the germination and growth of some weed species such as wild mustard (*Brassica* spp.) and poppy.

5. The integration of animals (e.g., cattle, swine, poultry) into farming systems, in addition to using them for milk, meat, and draft needs, adds another trophic level to the system, making it even more complex. Animals are fed crop residues and weeds with little negative impact on crop productivity, and this serves to turn otherwise unusable biomass into animal protein. Animals also recycle the nutrient content of plants by transforming them into manure. Furthermore, the need for animal feed broadens the crop base to include plant species that are useful for conserving soil and water (Reijntjes *et al.,* 1982). Legumes are often planted to provide quality forage, but they also improve the nitrogen content of soils. Integrated crop–livestock systems usually take the form of a crop–pasture rotation in which the pasture phase “charges” the system with nutrients and organic matter and the cropping phase “extracts” the accumulated nutrients. This balances biomass and nutrient inputs and outputs.

V. PRESERVING THE BIODIVERSITY OF TRADITIONAL AGROECOSYSTEMS

As many rural societies undergo the conversion from a subsistence economy to a cash agricultural economy, the loss of biodiversity in their ecosystems is mounting at an alarming rate. Because many peasants are directly linked to the market economy, external economic forces are increasingly influencing production by favoring genetically uniform crops and mechanized and/or agrochemical practices. Many landraces and wild plant relatives are being abandoned, which may cause them to become relic populations or even extinct. In some areas, land scarcity (mostly a result of uneven land distribution) has forced changes in land use and agricultural practices, which in turn have caused the disappearance of habitats that formerly maintained useful noncrop vegetation, including wild progenitors and weedy forms of crops (Altieri *et al.,* 1987).

In many parts of the world, genetic erosion is occurring at a fast pace because farmers are having to quickly change their farming systems because of economic, technical, and social pressures. As farmers adopt high-yield modern varieties (HYVs), they often subdivide their farming systems into commercial (mostly devoted to HYVs) and subsistence sectors, growing native varieties in the latter. The greatest loss of traditional plant varieties is occurring in lowland valleys close to urban centers and markets (Brush, 1986).

Given these destructive trends, many scientists and development workers have emphasized the need for *in situ* conservation of native crop genetic resources and the environments in which they occur (Prescott-Allen and Prescott-Allen, 1981). However, most researchers believe that *in situ* preservation of landraces would require a return to or the preservation of microcosms of traditional agricultural systems, which some regard as an unacceptable and impracticable proposition (Frankel and Soulé 1981). Nevertheless, the maintenance of traditional agroecosystems may be the only sensible strategy to preserve *in situ* repositories of crop germplasm. Although most traditional agroecosystems are undergoing some process of modernization or drastic modification, the conservation of crop genetic resources can still be integrated into agricultural development, especially in regions where rural development projects preserve the vegetation diversity of traditional agroecosystems and are anchored in the peasant rationale to utilize local resources and their intimate knowledge of the environment (Alcorn, 1984; Nabhan, 1983).

Previous recommendations for *in situ* conservation of crop germplasm emphasized the development of a system of village-level landrace custodians (a farmer curator system) whose purpose would be to continue growing a limited sample of endangered landraces native to the region (Mooney, 1983). One suggestion for preserving crop-plant diversity was for governments to set aside carefully chosen 5-by-20-km strips of land at as few as 100 sites around the world where native agriculture is still practiced (Wilkes and Wilkes, 1972). But given the increasing impoverishment and lack of income-generating alternatives for many rural populations in less developed countries, a proposition of this kind is clearly unrealistic since it fails to address the subsistence needs of these populations. In many areas where the urgent short-term goal of the local people is survival, diverting the limited land available for conservation purposes per se might prove totally inappropriate. A more feasible approach would be to support sustainable farming systems that incorporate native crops and wild/weedy relatives within and around production fields, as well as appropriate technologies aimed at upgrading food production for self-sufficiency (Altieri and Merrick, 1987). Such efforts would ensure that germplasm preservation remains linked to the economic and agricultural viability of local populations.

If biodiversity conservation is to succeed among small farmers, conservation goals and rural development efforts must be integrated to give equal importance to local resource conservation, food self-sufficiency, and equitable market participation. Any attempt at *in situ* crop genetic conservation must struggle to preserve the agroecosystem in which these resources occur (Nabhan, 1983). In the same vein, preservation of traditional agroecosystems cannot be achieved unless the sociocultural stability of the local community is also assured (Altieri, 1995).

An examination of effective grassroots rural development programs in less developed countries suggests that the process of agricultural improvement must (a) utilize and promote autochthonous knowledge and resource-efficient technologies, (b) emphasize use of local and indigenous resources, including valuable crop germplasm as well as essentials like firewood resources and medicinal plants, and (c) remain a self-contained, village-based effort with the active participation of the local people (Altieri, 1987). The subsidizing of a peasant agricultural system with external resources (e.g., pesticides, fertilizers, and irrigation water) can bring high levels of productivity, but such a system would then be sustainable only at high external cost and would depend on the uninterrupted availability of commercial inputs. In contrast, an agricultural strategy based on a diversity of plants and cropping systems can bring moderate to high levels of productivity through the manipulation and exploitation of the resources internal to the farm and can be sustainable at a much lower cost and for a longer period of time (Gliessman, 1998).

VI. USING BIODIVERSITY-BASED STRATEGIES TO SUPPORT TRADITIONAL AGRICULTURE

By understanding the common features of traditional agriculture, such as the capacity to bear risk, the use of biological folk taxonomies, and the production efficiencies derived from multiple and symbiotic crop mixtures, agricultural scientists have been able to develop technologies that support the needs and circumstances of specific groups. While subsistence farming generally lacks the potential for producing a meaningful marketable surplus, it does ensure food security. Many scientists wrongly believe that traditional systems do not produce more because hand tools and draft animals put a ceiling on productivity. However, where productivity is low, the cause appears to be social, not technical. When the subsistence farmer succeeds in providing food, there is no pressure to innovate or to enhance yields. Yet research shows that increased productivity is possible when traditional crop and animal combinations are adjusted and when labor and local resources are used more efficiently (Pretty, 1995).

As the inability of the Green Revolution to improve production and farm incomes for the very poor became apparent, growing enthusiasm for established, traditional agricultural practices generated a renewed quest in the developing world for affordable, productive, and ecologically sound technologies that could enhance small farm productivity while conserving resources. In the Andean altiplano, development workers and farmers have reconstructed a 3000-year-old indigenous farming system at an altitude of almost 4000 m. These indigenous farmers were able to produce food in the face of floods, droughts, and severe frosts by growing crops such as potatoes, quinoa, oca, and amaranthus in raised fields or “waru-warus,” which consisted of platforms of soil surrounded by ditches filled with water (Browder, 1989).

Technicians have now assisted local farmers in reconstructing 10 ha of these ancient farms, with encouraging results, which later led to a substantial expansion of the area under warus. For instance, yields of potatoes from waru-warus can surpass yields from chemically fertilized fields. Recent measurements indicate that waru-warus produce 10 tons of potatoes per hectare compared to the regional average of 1–4 tons/ha.

This combination of raised beds and canals has proven to have remarkably sophisticated environmental effects. During droughts, moisture from the canals slowly ascends the crop roots by capillary action, and during floods, furrows drain away excess runoff. Waru-warus also reduce the impact of temperature extremes. Water in the canal absorbs the sun’s heat by day and radiates it back by night, thereby helping protect crops from frost. On the raised beds, nighttime temperatures may be several degrees higher than in the surrounding area. The system also maintains its own soil fertility. In the canals, silt, sediment, algae, and organic residues decay into a nutrient-rich muck that can be dug out seasonally and added to the raised beds. There is no need for modern tools or fertilizers, and the main expense is manual labor to dig canals and build up the platforms. This ancient technology is proving so productive and inexpensive that it is now being actively promoted throughout the Andean altiplano.

One of the early projects advocating the reconstruction of traditional farming systems occurred in Mexico in the mid-1970s when the then existing Instituto Nacional de Investigaciones sobre los Recursos Bioticos (INIREB) unveiled a plan to build "chinampas" in the swampy region of Veracruz and Tabasco. Chinampa agriculture was perfected by the Aztec inhabitants of the Valley of Mexico prior to the Spanish Conquest. It involves the construction of raised farming beds in shallow lakes or marshes, and represents a self-sustaining system that has operated for centuries as one of the most intensive and productive ever devised by humans. Until the last several decades, chinampas demanded no significant capital inputs yet maintained extraordinarily high yields year after year. A wide variety of staple crops, vegetables, and flowers are often mixed with an array of fruit trees and bushes. Abundant aquatic life in the canals provides valuable sources of protein for local diets (Gliessman, 1998).

Now threatened by the sprawling growth of Mexico City and its suburbs, chinampas have nearly vanished except in a few isolated areas. Regardless, this system still offers a promising model as it promotes biological diversity, thrives without chemical inputs, and sustains year-round yields. When INIREB first began to establish the chinampa system in the lowland tropics of Tabasco, implementation and adoption met with mixed success. Some critics felt that no market outlets were explored or developed for the new outputs produced by the community. Nevertheless, the "raised beds" of Tabasco (or camellones chontales) are still in full operation in the swamps of this region, and apparently the local Chontal Indians have full control of them. The Chontal practice traditional agriculture, and these raised beds produce a great variety of products, which in turn have enhanced the income and food security of these "swamp farmers."

In a completely different ecoregion in the Andes, several institutions have engaged in programs to restore abandoned farming terraces and build new ones. In the Colca Valley of southern Peru, PRAVTIR (Programa de Acondicionamiento Territorial y Vivienda Rural) sponsors terrace reconstruction by offering peasant communities low-interest loans, seeds, and other inputs to restore large areas of abandoned terraces. The main advantages of using terraces are that they minimize risks in times of frost or drought, reduce soil loss, amplify the cropping options because of microclimate and hydraulic differences, and thus improve crop yields. Yield data from new bench terraces showed a 43–65 percent yield increase in potatoes, maize, and barley compared to yields of these crops grown on sloping fields. One of the main constraints of this technology is its high labor intensity, requiring about 350–500 worker-days per hectare for the initial building of the terraces. Such demands, however, can be buffered when communities organize and share tasks (Browder, 1989).

Another example of how a biodiversity-based approach can support or even resurrect traditional agriculture is occurring on Chiloé Island in southern Chile. This is a secondary center of origin of potatoes, and development workers are currently tapping the ethnobotanical knowledge of elderly female Huilliche Indians in an effort to slow genetic erosion and to recover some of the original native potato germplasm. They intend to provide impoverished farmers with locally adapted varieties that can produce without the use of agrochemical fertilizers. After surveying several agroecosystems on Chiloé, technicians collected hundreds of samples of native potatoes still grown by local farmers, and with this material, and in collaboration with farmers, they established community seed banks where more than 120 traditional varieties are grown year after year and are subjected to selection and seed enhancement. In this way, an *in situ* conservation program has been initiated involving farmers from various rural communities, thus ensuring the active exchange of varieties among participating farmers. As more farmers become involved, this strategy will provide a continuous supply of seeds to resource-poor farmers and will also create a repository of vital genetic diversity for future regional crop improvement programs (Altieri, 1995).

VII. CONCLUSIONS

A key conclusion that emerges from the relevant anthropological and ecological literature is that, when not disrupted by economic or political forces, indigenous modes of food production generally preserve rather than destroy biodiversity and natural resources. In fact, in any particular region, capitalist development through the promotion of large-scale, energy-intensive, commercial agriculture is bound to deplete natural resources more than some of the existing traditional systems. A number of studies have proven that many traditional agricultural systems are highly sustainable and productive, offering an alternative to the capital-intensive agriculture currently promoted by many development and governmental agencies. Besides employing crop diversity, traditional farmers use a set of practices that often cause minimal land degradation. These include the use of terraces and hedgerows in sloping areas, minimal tillage, mulching, small field sizes, and long fallow cycles (Grigg, 1974; Brush, 1982; Richards, 1985; Netting, 1993). It is clear that this more traditional strategy is both ecologically informed and environmentally sound, as the agricultural practices that are most likely to endure are those that deviate least from the native diversity of the natural plant communities within which they exist (Altieri, 1995; Gliessman, 1998).

This assessment of traditional subsistence agriculture does not romanticize its origins or practitioners, nor does it consider development per se to be detrimental. The intention is rather to stress the demonstrated value of traditional agriculture in the preservation of biodiversity, native crop diversity, and the adjacent vegetation communities (Toledo, 1980). Basing a rural development strategy on traditional farming and ethnobotanical knowledge not only assures the continual use and maintenance of valuable genetic resources, but also allows for the diversification of peasant or other indigenous subsistence strategies (Alcorn, 1984; Caballero and Mapes, 1985), which is a crucial issue in times of economic uncertainty.

The study of traditional agroecosystems and the ways in which indigenous peoples maintain and use biodiversity can facilitate the discovery of valuable agroecological principles, which in turn can contribute to the development of more sustainable agroecosystems and biodiversity conservation strategies in both developed and less developed countries.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • AGRICULTURE, SUSTAINABLE • ETHNOBIOLOGY AND ETHNOECOLOGY • GRAZING, EFFECTS OF • INDIGENOUS PEOPLES, BIODIVERSITY AND • LAND-USE PATTERNS, HISTORIC • TRADITIONAL CONSERVATION PRACTICES

Bibliography

Alcorn, J. B. (1984). *Huastec Mayan Ethnobotany*. University of Texas Press, Austin.

Altieri, M. A. (1987). *Agroecology: The Scientific Basis of Alternative Agriculture*. Westview Press, Boulder, Colorado.

Altieri, M. A., and M. K. Anderson. (1986). An ecological basis for the development of alternative agricultural systems for small farmers in the Third World. *Amer. J. Alternative Agriculture* **1**, 30–38.

Altieri, M. A., and L. C. Merrick. (1987). *In situ* conservation of crop genetic resources through maintenance of traditional farming systems. *Econ. Botany* **4**, 86–96.

Altieri, M. A., M. K. Anderson, and L. C. Merrick. (1987). Peasant agriculture and the conservation of crop and wild plant resources. *J. Soc. Conservation Biol.* **1**, 49–58.

Beets, W. C. (1982). *Multiple Cropping and Tropical Farming Systems*. Westview Press, Boulder, Colorado.

Berlin, B., D. E. Breedlove, and P. H. Raven. (1973). General principles of classification and nomenclature in folk biology. *Amer. Anthropologist* **75**, 214–242.

Brokensha, D. W., D. M. Warren, and O. Werner. (1980). *Indigenous Knowledge Systems and Development*. University Press of America, Lanham, Maryland.

Brush, S. B. (1982). The natural and human environment of the central Andes. *Mountain Res. Develop.* **2**, 14–38.

Brush, S. B. (1986). Genetic diversity and conservation in traditional farming systems. *J. Ethnobiol.* **6**, 151–167.

Bye, R. A. (1981). Quelites—Ethnoecology of edible greens—Past, present and future. *J. Ethnobiol.* **1**, 109–123.

Caballero, J. N., and C. Mapes. (1985). Gathering and subsistence patterns among the P'urhepecha Indians of Mexico. *J. Ethnobiol.* **5**, 31–47.

Chacon, J. C., and S. R. Gliessman. (1982). Use of the "non-weed" concept in traditional agroecosystems of southeastern Mexico. *Agro-Ecosystem* **8**, 1–11.

Chambers, R. (1983). *Rural Development: Putting the Last First*. Longman Group Limited, Essex, United Kingdom.

Chang, J. H. (1977). Tropical agriculture: Crop diversity and crop yields. *Econ. Geogr.* **53**, 241–254.

Clawson, D. L. (1985). Harvest security and intraspecific diversity in traditional tropical agriculture. *Econ. Botany* **39**, 56–67.

Denevan, W. M., J. M. Treace, J. B. Alcorn, C. Padoch, J. Denslow, and S. T. Paitan. (1984). Indigenous agroforestry in the Peruvian Amazon: Bora Indian management of swidden fallows. *Interciencia* **9**, 346–357.

Francis, C. A. (1986). Variety development for multiple cropping systems. *CRC Crit. Rev. Plant Sci.* **3**, 133–168.

Frankel, O. H., and M. E. Soulé. (1981). *Conservation and Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Gliessman, S. A., E. Garcia, and A. Amador. (1981). The ecological basis for the application of traditional agricultural technology in the management of tropical agro-ecosystems. *Agro-Ecosystems* **7**, 173–185.

Grigg, D. B. (1974). *The Agricultural Systems of the World: An Evolutionary Approach*. Cambridge University Press, Cambridge, United Kingdom.

Harlan, J. R. (1976). The possible role of weed races in the evolution of cultivated plants. *Euphytica* **14**, 173–176.

Harwood, R. R. (1979). *Small Farm Development—Understanding and Improving Farming Systems in the Humid Tropics*. Westview Press, Boulder, Colorado.

Klee, G. A. (1980). *World Systems of Traditional Resource Management*. J. Wiley & Sons, New York.

Mooney, P. R. (1983). The law of the seed. *Develop. Dialogue* **1**, 1–172.

Nabhan, G. P. (1983). Papago Indian Fields: Arid Lands Ethnobotany and Agricultural Ecology. Unpubl. Ph.D. diss., University of Arizona, Tucson.

Norman, M. J. T. (1979). *Annual Cropping Systems in the Tropics*. University Presses of Florida, Gainesville.

Prescott-Allen, R., and C. Prescott-Allen. (1981). In Situ Conservation of Crop Genetic Resources: A Report to the International Board for Plant Genetic Resources. IBPGR, Rome.

Richards, P. (1985). *Indigenous Agricultural Revolution*. Westview Press, Boulder, Colorado.

Toledo, V. M. (1980). La ecologia del modo campesino de producion. *Antropologia y Marxismo* **3**, 35–55.

Toledo, V. M., J. Carabias, C. Mapes, and C. Toledo. (1985). *Ecologia y Autosuficiencia Alimentaria*. Siglo Vientiumo Editors, Mexico City.

Wilken, G. C. (1970). The ecology of gathering in a Mexican farming region. *Econ. Botany* **24**, 206–245.

Wilken, G. C. (1977). Integrating forest and small-scale farm systems in middle America. *Agro-Ecosystems* **3**, 291–302.

Wilkes, H. G., and K. K. Wilkes. (1972). The green revolution. *Environment* **14**, 32–39.

Williams, B. J., and C. Ortiz-Solario. (1981). Middle American folk soil taxonomy. *Ann. Assoc. Amer. Geographers* **71**, 335–358.

AIR POLLUTION

Mike Ashmore
University of Bradford

GLOSSARY

acute toxicity A damaging effect caused by a single short period of exposure to high concentrations of a pollutant.

available (or bioavailable) concentration The concentration of a contaminant in soils or surface waters which can be taken up by the target organism.

background concentration The concentration of a contaminant in an environment which has not been measurably influenced by anthropogenic sources.

chronic toxicity A damaging effect caused by a long period of exposure to low or moderate concentrations of a pollutant.

contamination The presence of elevated concentrations of a toxic substance, compared with normal ambient concentrations.

deposition The rate of influx of a substance from the atmosphere, usually expressed as mass per unit area of ground.

detoxification Reduction in the toxic effect of a pollutant by its chemical or biological transformation to another less toxic chemical.

emission The release of a contaminant to the environment.

photochemical Chemical reactions with rates which are increased by radiation of particular wavelengths.

phytotoxic Damaging to plants.

pollution The presence of a toxic substance at a concentration that can cause adverse effects.

wet deposition The influx of contaminants from the atmosphere in precipitation (i.e., rain, snow, mist, fog, and clouds).

AIR POLLUTION results from a range of sources, including industry, power generation, transport, and domestic sources. Its local impacts on biodiversity can be dramatic close to large point sources of emissions, but significant impacts on biodiversity over much wider areas can also result from the long-range transport of pollutants. This article provides a brief overview of the mechanisms of air pollution impacts on biodiversity and of the evidence of such impacts in the field. It also considers the issues connected with establishing credible air quality guidelines to protect biodiversity.

I. SOURCES AND ENVIRONMENTAL PATHWAYS OF AIR POLLUTANTS

A. Historical Overview

The impact of air pollutants derived from human activity on vegetation and animals has been recognized for many centuries. The English diarist John Evelyn wrote a pamphlet on air pollution in London, published in 1661, in which he noted, "It is the horrid smoake which kills our Bees and Flowers abroad, suffering nothing in our gardens to bud, display themselves or ripe." There is also evidence from historical records of changes in biodiversity as a result of pollution. For example, until the early 1800s the dominant vegetation of the upland bogs of the Peak District in the southern Pennines of England was *Sphagnum* moss species. However, with the onset of the industrial revolution in the surrounding valleys, pollutant levels in the region increased rapidly. The *Sphagnum* species began to disappear, and today the area is dominated by cotton grass (Eriophorum vaginatum). Overall, there has been a loss of species diversity as species associated with the ombrotrophic Sphagnum species, such as sundews (*Drosera* species) and marsh andromeda (*Andromeda polifolia*), have also been lost. There is little doubt in this case that the change in community composition is a result of the direct elimination of the sensitive *Sphagnum* species by increased sulfur deposition.

During the past century, the range and quantities of chemicals discharged into the atmosphere from industry, transport, agriculture, energy production, domestic heating, and many other human activities have increased dramatically. Once discharged into the atmosphere, these compounds are physically dispersed in the atmosphere and may undergo chemical transformations that alter their potential environmental impact. The environmental impact of air pollutants will depend on their concentration in the environment or on the dose received or accumulated by the target organism. At low doses, the environment may be contaminated but no adverse effects ensue; normally, the chemical is only referred to as a pollutant if it has some environmental impact.

B. Sources, Distribution, and Effects of Major Air Pollutants

Air pollutant problems vary greatly in their spatial scales. Some are very local in character, with the environmental impact of the pollution restricted to the immediate vicinity of, for example, a road or a factory. Other problems are regional in character as a result of the long-range transport of pollutants such as acid rain and tropospheric ozone. Similarly, pollutant impacts may vary on different temporal scales. Some impacts, for example, are the result of an accidental release of large pollutant concentrations, which may cause an immediate impact on biodiversity, and from which there may be a slow and gradual recovery, whereas others are the result of an accumulation of pollutant deposited over years or even decades.

This chapter provides an overview of the ways in which pollutants can affect ecological processes and

TABLE I

Summary of Major Sources and Impacts of Air Pollutants of Relevance to Biodiversity

Pollutant	Major sources	Major impacts	Scale of effects
Sulfur dioxide (SO_2)	Power generation; industry; domestic and commercial heating	Forest decline; elimination of lichens and bryophytes	Local
Nitrogen oxides (NO_x) and ammonia (NH_3)	Power generation and transport (NO_x); intensive agriculture (NH_3)	Altered plant growth and enhanced stress sensitivity; soil acidification and eutrophication	Local, regional
Acid deposition	Secondary pollutant formed from SO_2 and NO_x	Soil and freshwater acidification; forest decline	Regional
Ozone (O_3)	Secondary pollutant formed from hydrocarbons and NO_x	Reduced plant growth; forest decline	Regional
Toxic metals (e.g., lead and cadmium)	Smelting industry; transport (lead)	Reduced soil microbial activity; reduced soil invertebrate populations	Local, regional
Persistent organic pollutants (POPs)	Industry; fuel combustion; pesticide use	Bioaccumulation in food chain	Local, global

biodiversity. Inevitably, it is not possible to provide a comprehensive account of the effects of the vast range of contaminants emitted into the atmosphere by human activity; therefore, this chapter will concentrate on those air pollutants for which the greatest evidence exists of impacts on biodiversity. Table I summarizes the major sources of these pollutants, their major ecological impacts, and the spatial scale of their impacts.

C. Deposition and Environmental Pathways

The pathways by which pollutants enter ecosystems may have direct relevance to the nature of their impacts. Where pollutants enter as gases, they may be taken up directly into foliage, through stomata, or directly inhaled by animals. However, in other cases pollutants may enter ecosystems in rainfall, in occult precipitation (cloud or mist, in which pollutant concentrations are often higher than in rain), or as particles.

The pathways of aerially deposited pollutants through ecosystems and the mechanisms by which they affect biodiversity are varied. In general, there are four groups of air pollutants which are of significant concern in terms of biodiversity and which differ generically in the nature of their pathways from the atmosphere to the site of biological impact:

1. Acidifying and eutrophifying pollutants, such as sulfur dioxide, nitrogen oxides, and ammonia, which may also be deposited as sulfate, nitrate, and ammonium in wet deposition: These gases may have direct effects on organisms, but over longer periods they can also lead to acidification of soils and fresh waters and to changes in their nutrient status, with implications for biodiversity.
2. Photochemical oxidants, primarily ozone: These compounds are not emitted directly into the atmosphere but are secondary pollutants formed as a result of reactions involving nitrogen oxides and volatile organic compounds. These reactions require sunlight and high temperatures, and hence photochemical oxidants are characteristic of warmer climates. The main impacts of ozone on biodiversity are through direct uptake into, and effects on, leaf tissue.
3. Metals and other inorganic contaminants, which are deposited primarily in rainfall or in particulate matter: Atmospheric deposition is only one source of these chemicals; they can also result from soil contamination or from discharges into marine or freshwater systems. The impacts of metals result primarily from their accumulation in soils at concentrations which are toxic to soil organisms or plant roots or through leaching into fresh waters.
4. Persistent organic pollutants (POPs): These are of concern because of the potential for significant bioaccumulation in the food chain. As for metals, emissions to the atmosphere are not the only source of these chemicals. However, unlike for metals, there is evidence that the atmosphere can act as a pathway to redistribute these compounds through "global distillation," in which compounds volatilize at ambient temperatures in warmer parts of the planet and are then redeposited at cooler latitudes. Thus, it is the bioaccumulation of these compounds in the polar regions which is of greatest concern; for example, Antarctic fish have been reported to contain concentrations of certain POPs which are as high as those in North Sea fish.

D. Current and Future Global Distribution of Air Pollution Emissions

In the past few decades, measures have been taken in Europe and North America which have reduced atmospheric emissions of many pollutants. Although recovery of biodiversity in ecosystems affected by these pollutants may take time, and may not result in the same community as that originally present in a particular location, major new impacts on biodiversity due to large increases in atmospheric emissions are unlikely in these regions. In contrast, the past two decades have seen large increases in atmospheric emissions in parts of Asia, Africa, and Latin America, and continued increases are projected. It is estimated that sulfur dioxide (SO_2) emissions in 2010 will be higher in Asia than the combined emissions of Europe and North America. These projections are based on projected growth in population and GDP and assume that no significant measures to reduce emissions will be taken in Asia, Africa, or Latin America. Similar changes in the global patterns of emissions of nitrogen oxides have been projected for the first decades of this century.

II. MECHANISMS OF IMPACTS ON BIODIVERSITY

A. Long-Term and Short-Term Effects on Individual Organisms

Before considering the impacts of air pollution at a community level, it is essential to assess the ways in which individual plants and animals respond to air pollutants. The impact of a pollutant on any individual organism is complex and may involve many factors. The most important of these factors is the dose of the pollutant received. This will partly depend on the concentration of the pollutant in the atmosphere, or other relevant medium, and the duration of exposure to it. Short-term effects of air pollution exposure at high concentrations result in acute toxicity which is usually characterized by direct damage to exposed tissue and visible foliar injury on vegetation. In contrast, long-term effects of air pollution exposure, which may result from much lower concentrations, result in chronic toxicity, which is usually characterized by alterations in physiology, growth, and reproduction.

However, pollutants may be present in the environment without actually being taken up by the organism. For example, when stomata are closed at night, or under drought conditions, less air pollution can penetrate to the internal leaf tissue, although pollutants may still be deposited on the leaf surface. Similarly, the impact of metals deposited from the atmosphere on soil microorganisms is believed to relate most closely to the free ion concentration in soil solution; for metals such as lead, which are tightly bound to the soil matrix, this may only be a small fraction (the bioavailable fraction) of the total soil concentration.

In general, the relationship between pollutant concentration, or dose, and the response of an organism is classified as follows:

1. When the pollutant is present in sufficiently high concentrations, the organism may be killed outright.
2. At lower concentrations, the organism is able to survive, but its performance is adversely affected. For example, rates of growth or reproduction may be reduced, or changes may occur in the patterns of development or in resource allocation.
3. At even lower concentrations, the physiology and growth of the organism may be unaffected under optimal environmental conditions, but the pollutant may cause subtle morphological, physiological, or behavioral changes which lead to altered tolerance of other environmental stresses.
4. Finally, in the case of certain pollutants, there may be positive effects at low concentrations. For example, certain toxic metals, such as copper and zinc, are essential micronutrients in animal diets, whereas sulfur dioxide can stimulate plant growth at low concentrations, especially in soils which are sulfur deficient.

This range of different biological responses, depending on the severity of the air pollution stress, means that different types of mechanisms exist for impacts on biodiversity. Thus, for 1, direct elimination of species may lead directly to a reduction in biodiversity. For 2, reductions in biomass and population growth rates may affect the longer term viability of populations and lead to local extinctions; however a more likely outcome is a shift in competitive balance between species, which may lead to a changed community composition. In the case of 3, the pollutant has no direct effect on biodiversity but may exacerbate, or mitigate, the impacts of other stress factors. Finally, for 4, although the effect of the air pollutant may superficially be positive, increases in growth of one species may alter the competitive balance and lead to reduced populations or local extinctions of other species. Air pollution may also have effects on some species indirectly through its impacts on other components of the ecosystem. For example, elimination of forest cover through direct effects of air pollution will cause large changes in the microclimate of the forest floor, with consequences for ground vegetation and soil biodiversity.

There is a large degree of variation between species in their sensitivity to particular pollutants; there is also often substantial genetic variation in response within species. This variation broadly relates to the ability of the organism to restrict pollutant uptake or, once it has been taken up, to detoxify, metabolize, or sequester the pollutant. Within the same genotype, other factors such as age and growth stage may also influence sensitivity to air pollutants. Finally, it is important to realize the dynamics of this response, with pollutant exposure frequently inducing adaptive biochemical, physiological, or morphological responses which lead to a reduction in its adverse effects.

B. Interactions between Air Pollutants and Other Environmental Factors

The effect of a given dose of an air pollutant may also depend on other environmental conditions which can modify responses to the pollutant in many ways. First, they may modify pollutant dose; for example, soil water stress may lead to stomatal closure and thus to a reduced uptake of air pollutants by plants and hence reduced pollutant damage. Second, environmental factors may reduce the capacity of an organism to detoxify and assimilate pollutants; for example, it is well established, from both controlled experiments and field observations, that SO_2 is more phytotoxic when plants are growing at low temperatures or under low light conditions. Third, exposure to pollutants may lead to changes in the morphology and/or physiology of the organism which make it more sensitive to environmental stresses; for example, increased deposition of sulfate or ammonium ions to conifer seedlings has been shown to increase their sensitivity to cold stress.

Air pollutants rarely occur alone, and the responses to pollutant mixtures may be very different from those to the individual constituent pollutants. These interactions can operate in several different ways. First, two pollutants taken up together may have a greater effect than would be expected from knowledge of the effects of each pollutant—a so-called synergistic response. For example, inhalation of SO_2 and particles together often has a synergistic effect on humans and animals, whereas uptake of SO_2 and nitrogen dioxide (NO_2) together often has synergistic effects on vegetation. Although synergistic interactions between air pollutants have received considerable attention, there are many instances in which the effect of a pollutant mixture is not different from that of the individual pollutants or indeed there is a reduced effect—an antagonistic interaction. Second, over the longer-term, deposition of one pollutant can affect the uptake and impacts of another contaminant; for example, there is evidence that freshwater acidification resulting from deposition of sulfate and nitrate can cause increased bioaccumulation of metals such as lead, cadmium, and methylmercury in fish and birds. Finally, problems such as forest decline may result from complex interactions between a whole range of air pollutants, biotic and abiotic stresses, and management factors.

C. Evolutionary Responses to Air Pollution Exposure

In competitive situations, it is likely that selection pressures will act in favor of individuals that are more resistant to a particular stress factor, and there is no reason that air pollution should be an exception to this, given the established intraspecific variation in response to pollution. There has long been an awareness among entomologists of these effects of air pollution, as exemplified by the evolution of industrial melanism in moths and ladybirds. Darker forms of these species have a selective advantage where trees have a darkened bark and a reduced lichen cover caused by high concentrations of SO_2 and smoke because of reduced predation. Improvements in air quality have led to a reversal of this evolutionary trend, with a decreased proportion of darker forms in the population.

Evolution of tolerance to air pollution has also been demonstrated in vegetation. Studies of emission sources in both the United States and the United Kingdom have shown that tolerance to SO_2 in annual and perennial species increases with increasing concentrations closer to the source. There is also evidence that decreasing urban SO_2 concentrations are associated with a loss of this tolerance in grass species, suggesting that the SO_2-tolerant genotypes are at a selective disadvantage in the absence of the pollutant. Where the pollutant is more widely dispersed, it is more difficult to demonstrate the phenomenon, although there is no reason to suppose that it does not occur. Thus, recent work in the United Kingdom has shown both temporal and spatial associations between ozone exposures and the ozone tolerance of local populations of the annual species *Plantago major* which are strongly suggestive of an evolution of tolerance.

There is little evidence of the impact of evolution of air pollution tolerance on the overall genetic variation, or fitness, of a population. Key factors are the strength of the selective pressure from air pollution and the frequency of resistant genotypes. Where pollution stress is high and the frequency of resistant individuals is low, and a large proportion of the population is eliminated, the effect on overall genetic variation in the population may be substantial. However, with less severe pollution levels, the effect on genetic variation may be much less.

D. Effects on Interactions between Organisms

When an air pollutant is present at a concentration which affects the physiology, growth, or reproduction of individual organisms, it is clear that the potential exists to influence the outcome of competition between species. In the case of competition between plant species, several simple experiments with air pollutants such SO_2 and ozone (O_3) have shown, as expected, that when a pollution-sensitive and pollution-resistant species (often a clover and grass species, respectively) are grown together, the presence of the pollutant shifts the balance in favor of the latter. However, although these simple experiments clearly demonstrate that pollutants can modify the outcome of plant competition, it is doubtful whether they provide much guidance regarding how pollution modifies interactions between plant species in real communities. For example, the vertical stratification of the plant community, which cannot be readily reproduced in simple competition experiments, may have a major effect on community responses. Field studies of forest stands which have been damaged by industrial pollution typically report a decline in density of the dominant overstory tree species. However, in some cases this can be accompanied by an increase in density in the lower canopy and in the shrub layer as a result of the release of competitive suppression.

Air pollution is also known to alter the relationships between plants and insect herbivores. For example, high concentrations of nitrogen oxides are probably a major factor in the high numbers of aphids, and other phytophagous insects, found alongside major roads. Figure 1 illustrates some of the potential interactions between air pollution, insect pests, and host plants. At high concentrations, there may be direct effects on the insects, but at lower concentrations the effect is mediated primarily through chemical changes in the host plant; for SO_2 and NO_2, the effects on amino acid composition may be particularly important. Effects on the natural enemies of insect herbivores are another possible mechanism of response. These primary interactions can, as Fig. 1 demonstrates, lead to a range of secondary interactions affecting both plant performance and populations of other species at higher trophic levels.

Air pollutants can similarly modify the interactions between plants and fungal pathogens. Many pathogens are very sensitive to pollutants such as SO_2, and the absence of certain diseases producing clear visible symptoms, such as tarspot on sycamore (*Rhytisma acerinum*), has been proposed for use as a bioindicator of elevated concentrations of SO_2. Chemical and biological

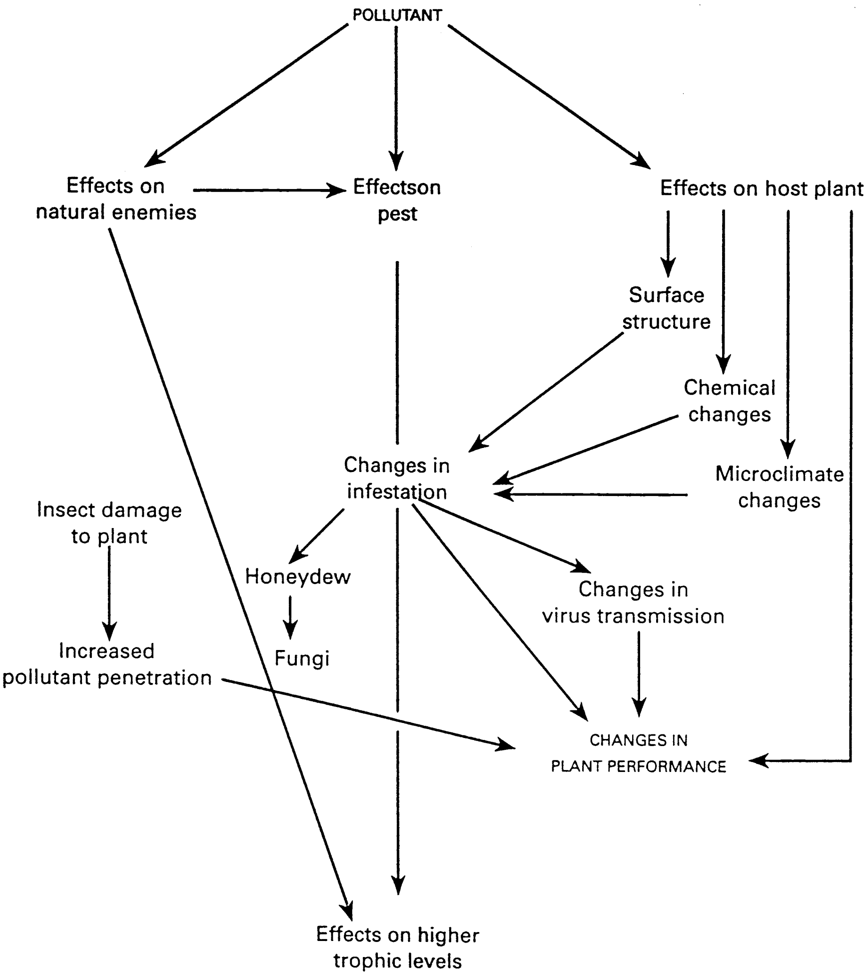


FIGURE 1 Model of potential interactions between air pollution, insect herbivores, and host plants [source: Bell, J. N. B., *et al.* (1993). *Parasitology* **106**, S11–S24, Cambridge University Press].

changes on the leaf surface induced by pollutants may critically affect the performance of fungal pathogens and other leaf microflora. Similarly, acid deposition can lead to acidification of tree bark, affecting the range of epiphytic lichen species found.

Pollutant exposure can have an adverse effect on mycorrhizal associations, which are crucial to the stress tolerance (including heavy metal tolerance) and competitive ability of many plant species. This may occur because of direct effects of changes in soil chemistry on some mycorrhizal types due, for example, to mobilization of aluminum by acid deposition or accumulation of heavy metals. Alternatively, exposure to air pollutants above-ground may influence mycorrhizal activity indirectly through changes in carbon partitioning to the roots.

III. EVIDENCE OF IMPACTS ON BIODIVERSITY

The previous section summarized the key factors which may influence the impact of air pollution on biodiversity and provided some simple examples. In this section, some examples of actual impacts of air pollution on biodiversity in the field are discussed in more detail to illustrate the types of effects which have been found. It should be noted that the evidence proving the link between exposure to air pollution and loss of biodiversity is variable in quality. Since experimental manipulations of actual communities over the timescales of decades over which these effects may develop are impossible, it is generally necessary to rely on field evidence. Where the effects are severe and localized, spatial associations between pollutant deposition and biodiversity may be readily established and, especially when the species present before operation of a local source are known, a casual link may readily be inferred. However, where the concern is related to regional-scale deposition of air pollutants at more moderate levels and the effects on biodiversity are more subtle, possibly involving interactions between several factors, the causal link is much more difficult to establish. In such cases, simple cause–effect relationships do not exist, and we are reliant on establishing a balance of probabilities, or using a precautionary approach, in establishing the need for intervention to reduce pollutant emissions to protect biodiversity. The examples discussed in the following sections illustrate a range of cases, from local to regional.

A. Effects of SO_2 and Metals Near Large Smelters

It is clear that when air pollutants are present in high enough concentrations they can cause the complete elimination of all plant species. Near the large Sudbury smelter in Ontario, for example, in the early 1970s areas devoid of vegetation occurred up to 8 km from the source, and species numbers and productivity were reduced up to 20–30 km from the smelter. Figure 2 illustrates the relationships between cover, species numbers, diversity index of overstory vegetation, and the distance from the smelter; similar relationships were found for ground vegetation, although the trend with distance was not as strong. These effects were primarily due to the combined effects of SO_2 and heavy metal emissions from the smelter. There was also evidence of acidification and heavy metal accumulation in lakes near the smelter, with adverse effects on the numbers of species of plankton, macrophytes, and fish. During the past 20 years, there has been a reduction in emissions of both SO_2 and heavy metals from the smelter, and some reinvasion is evident, particularly of metal-tolerant ecotypes of certain grass species.

Similar patterns of decreases in species diversity and productivity have been reported near other large point sources of SO_2 and metals. It is of interest to compare the results found near these sources, which, unlike Sudbury, have minimal SO_2 emissions. One of most intensively studied areas is near the brassworks at Gusum, Sweden, where metals have accumulated in the surrounding soil over a period of three centuries. Figure 3 shows the relationship between soil copper concentrations and the numbers of taxa found for different groups of organisms; it should be noted that lead and zinc are associated with copper, and it should not be assumed that copper is the sole causal factor. Vascular plant species are relatively unaffected in terms of numbers of taxa, although cover and growth are reduced near the smelter; in contrast, mosses, earthworms, ground lichens, and macrofungi are all significantly reduced in terms of numbers of taxa at higher metal concentration.

Patterns of recovery as emissions decrease reflect the pathways of the pollutants. For example, a series of studies in woodlands near a major metal smelter in southwest England revealed that most of the deposited metal accumulated in the soil compartment, but that there are contrasting patterns of concentration change in the soil profile in response to decreasing emissions for four metals. In the case of cadmium and zinc, concentrations in the surface horizons have decreased since

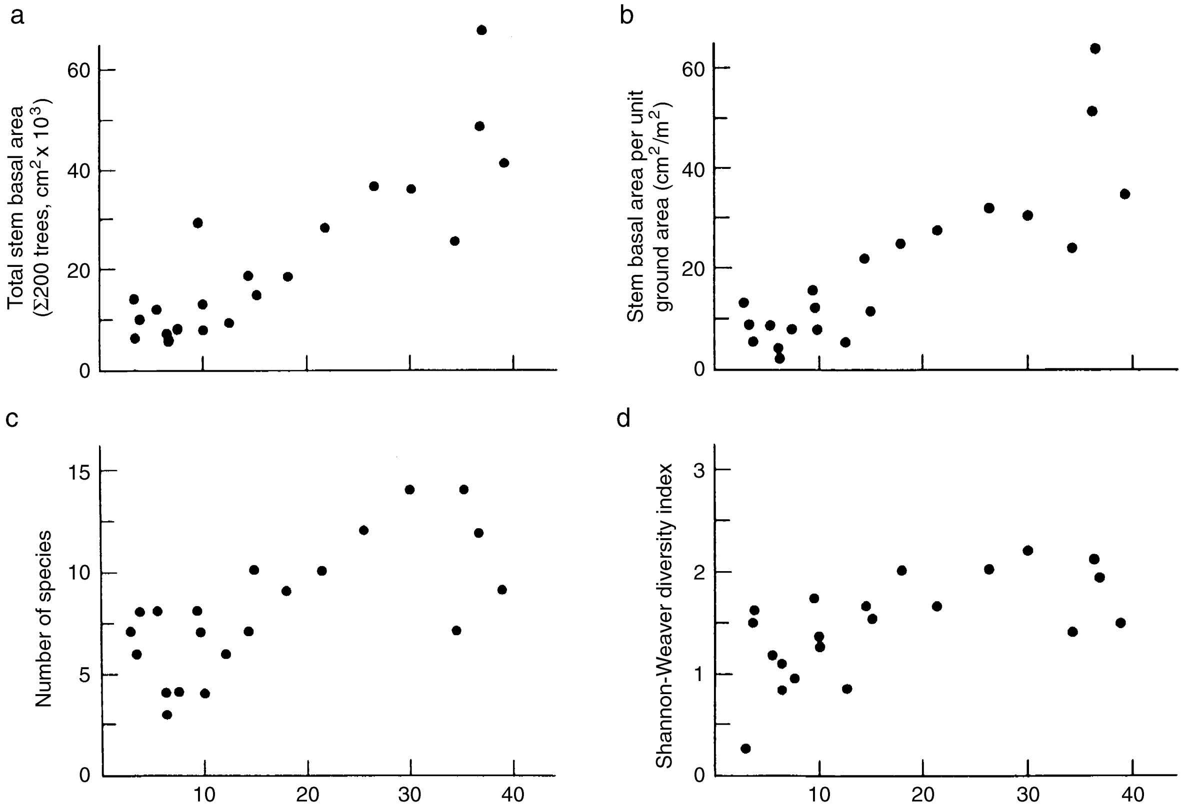


FIGURE 2 Forest overstory characteristics, expressed as total stem basal area (a), per unit ground area (b), number of species (c), and Shannon–Weaver diversity index, with distance south of the Sudbury smelter complex [source: Freedman (1995); redrawn with permission from Freedman, B., and Hutchinson, T. C. (1980). *Can. J. Bot.* 58, 2123–2140].

the mid-1970s and there is evidence of a wave of higher concentrations passing down the soil profile. In contrast, the concentrations of lead and copper, which are highly bound to the soil matrix, have continued to increase in the surface layers, and there is no evidence of movement down the profile. The impact of smelters also depends on soil and climate. For example, the impact of two large nickel smelters in the Kola peninsula, which produce large emissions of metals and are estimated to be major European sources of SO_2, may be exacerbated by their location in the Arctic, where nutrient turnover rates are low and the low temperatures are well documented to increase the sensitivity of vegetation to SO_2.

B. Effects of Air Pollution on Lichen Biodiversity

The impacts of air pollution, and SO_2 in particular, on lichen species provide another illustration of large-scale changes in biodiversity. The disappearance of many lichen species from European and North American cities during the past century has been well documented; in the United Kingdom, between 30 and 90% of species were lost from areas in which air pollution was a dominant factor. Many of these species were highly sensitive to direct effects of relatively low concentrations of SO_2. In contrast, other species are relatively tolerant, and several different scales have been developed to map SO_2 concentrations based on the occurrence of particular groups of lichen species.

In many of these cities, SO_2 concentrations have decreased dramatically in recent years, but the reinvasion of lichens in response to this decline has been patchy and variable. In London, for example, it has been found that some relatively pollution-sensitive lichens have reinvaded more quickly than species which are more tolerant. This variable pattern of recovery means that the bioindicator scales based on lichen species distributions can no longer provide a reliable basis

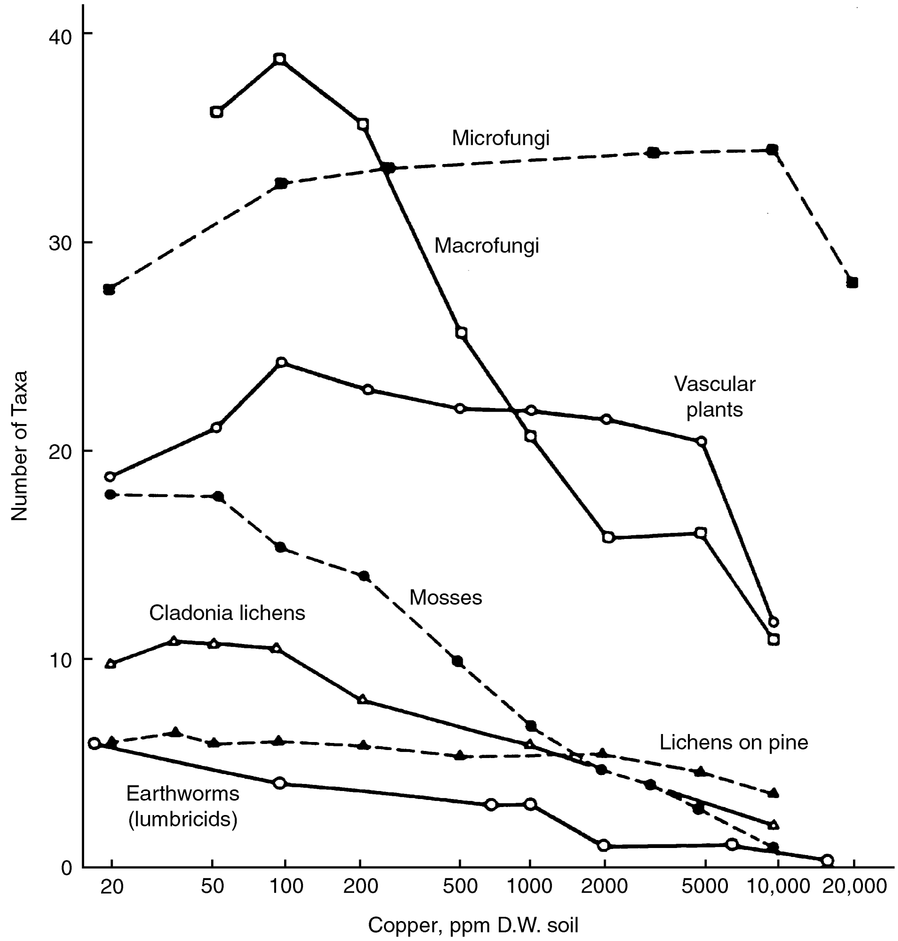


FIGURE 3 Species richness of different groups of organisms as related to copper concentrations in the surface soil of forests near the Gusum brassworks [source: Freedman (1995); redrawn with permission from Tyler, G. (1994). *Ambio* **13**, 18–24].

for mapping SO_2 concentrations; the lichen communities are no longer in equilibrium with the pollutant levels. There are several possible explanations for these different rates of recovery. One factor is certainly the speed of dispersal of vegetative propagules into areas of decreasing SO_2; certain species (so-called "zone-skippers") are able to disperse much more rapidly than others. Another important factor is changes in bark chemistry: Recovery of epiphytic species seems to be more rapid on tree species such as ash and willow, which have a relatively high pH bark, than on species with acidic bark, such as oaks. Once bark has been acidified, there may be a delayed recovery of the lichen flora as SO_2 concentrations decrease because bark chemistry changes more slowly than atmospheric concentrations.

There is also evidence of other pollutants affecting the wider distribution of epiphytic lichens. Acid deposition in more remote areas of the United Kingdom and Sweden has been linked to the loss of sensitive cyanobacterial lichens, such as certain *Lobaria* species; as in the urban situation, these effects are more marked on tree species with acidic, poorly buffered bark. Increased levels of ammonium deposition have also been associated with an increase in the cover and distribution of nitrophytic species and a loss of more acidophytic species in The Netherlands. Thus, although there is little doubt from field evidence that there are currently considerable changes in lichen distributions in many areas of western Europe, a complex mixture of causal factors is involved that includes interactions between changing patterns of pollutant deposition, substrate chemistry, dispersal mechanisms, and other factors.

C. Effects of Nitrogen Deposition on Heathland Communities

A recent example of changes in biodiversity which may be linked to atmospheric pollution, but are not due to direct adverse effects of these pollutants, comes from The Netherlands. Here, in recent decades, heathlands dominated by ericaceous shrubs, such as *Calluna vulgaris*, have been replaced in many areas by acid grassland communities dominated by *Molinia caerulea* and *Deschampsia flexuosa*. A major cause of this change in community structure has been identified as increased nitrogen deposition due primarily to increased emissions of ammonia from intensive agriculture.

Here, the primary mechanism postulated is a shift in the competitive balance between *Calluna* and the grass species; experimental studies clearly demonstrate that young plants of the grass species are better able to respond to increased levels of nitrogen and thus can out-compete *Calluna*. However, the field situation is more complex because of the effect of competition for light; once a *Calluna* canopy is established, invasion by grasses is unlikely, even when there are high levels of nitrogen availability. Thus, it has been suggested that the increased nitrogen deposition also acts by increasing the sensitivity of *Calluna* to biotic and abiotic stress factors, which will lead to canopy breakdown; in particular, outbreaks of heather beetle infestations, which are favored by higher leaf nitrogen concentrations, can cause widespread defoliation of the *Calluna* canopy. Finally, it is likely that management practices are also significant; there have been reduced rates of grazing and sod-cutting in recent years, and these traditional management practices have been important in maintaining the low nutrient status of the heathland communities. Thus, in this case, changes in species composition may have resulted from a complex mixture of factors, including pollutant deposition, management practices, competition for light and nutrients, and the effects of climatic stress and insect herbivores.

D. Effects of Ozone on Forest Ecosystems

The three previous cases involved strong spatial gradients in air pollution concentrations due to industrial, urban, and agricultural sources of pollution. In each case, the major pollutants responsible for the observed effects were primary pollutants. In contrast, the secondary pollutants involved in the final two cases are more regional in their distribution, and the spatial and temporal gradients in pollutant exposure are often confounded by other factors.

Where these pollutants are present in high concentrations, the evidence linking cause and effect is strong. For example, the area where the impacts of ozone stress on forest community composition has been most intensively studied is in the San Bernadino mountains, which surround the city of Los Angeles. Effects of ozone pollution, generated from pollutant emissions in the city, began to be observed on the native forest community in the 1960s. The most dominant species of these mixed-conifer forests prior to European settlement were ponderosa pine (*Pinus ponderosa*) and Jeffrey pine (*Pinus jeffreyi*) because of their tolerance of the frequent wildfires; however, these have also proved to be the most sensitive species to ozone. In many of these areas, both species have shown severe foliar injury and reduced needle longevity. These are associated with reduced radial growth or even years with missing growth rings. Trees affected by ozone are more susceptible to attack by bark beetles, which are often the direct cause of mortality. Outbreaks of bark beetles are associated with drought years and high ozone concentrations. Regeneration in these forests is greater for trees such as white fir or cedar species, which are more resistant to ozone, although at some higher elevation sites these and other conifer species do not naturally regenerate and the area may become dominated by shrubs. The patterns of change in community composition are confounded by the role of fire since current fire exclusion policies favor replacement of ponderosa and Jeffrey pine by more fire-sensitive species which also happen to be more ozone tolerant.

Experimental studies of the effects of ozone on other tree, shrub, and herbaceous species have clearly demonstrated that, when grown alone, many of these species show significant effects of ozone at concentrations which are found in the field. Systematic surveys of visible leaf injury in the United States have clearly demonstrated the presence of such injury on many native forest and herbaceous species, but the significance of such symptoms for long-term effects on biodiversity is unclear. For example, a recent ozone fumigation study of an early successional forest community showed that ozone decreased species richness, diversity, and evenness; however, the species which dominated the community in the highest ozone treatment was blackberry, which is considered to be highly ozone sensitive on the basis of visible injury responses.

The previous examples serve to illustrate that, even where pollutant concentrations are high enough to cause significant levels of visible leaf injury, the short-term and long-term effects of ozone on canopy composition may be complex and not readily predictable because of the interactions with management practices, biotic and abiotic stresses, and competition between species. Thus, assessment of the long-term effects in other forest ecosystems in which declines of sensitive species have been attributed to the effects of ozone, such as the declines in fir vitality which have been reported recently in the mountains near Mexico City, will need to take account of local factors which may significantly modify responses to the pollutant.

E. Effects of Sulfur and Nitrogen Deposition on Forest Ecosystems

The long-term effects of ozone on biodiversity are difficult to predict because we lack an understanding of the key mechanisms involved at an ecosystem level. In contrast, the effects of deposition of sulfate, nitrate, and ammonium can be understood more mechanistically in terms of their effects in causing soil acidification or eutrophication and in terms of nutrient cycling. Thus, any analysis of the causes of any change in forest vitality needs to consider atmospheric, edaphic, and biological factors influencing nutrient cycling as well as their interaction with other stress factors and with forest management. Figure 4 illustrates the key interactions between these factors. In the case of ozone, the analysis of impacts on forest ecosystems is focused on the direct effects of the air pollutant on the forest canopy, as modified by site factors, pests and diseases, and climate. In contrast, when considering the impacts of sulfur and nitrogen deposition, direct effects on the forest canopy are of minor concern except at very high rates of deposition. Rather, it is the long-term effects on soil, soil solution, and soil leachate chemistry, and their implications for plant growth and vitality, which are crucial.

In soils in which buffering is dominated by cation exchange, increased acid deposition may lead to exchange and leaching of base cations, such as calcium and magnesium, down the soil profile and beyond the rooting zone. In more acidic soils, buffering is dominated by aluminum exchange, and acid deposition can increase levels of available aluminum, which can be directly toxic to fine roots or restrict uptake of other mineral nutrients. Soil acidification may also increase the availability of other toxic metals and can have a direct impact on rates of litter decomposition. There is strong evidence of acidification and base cation depletion of forest soils in western Europe during the past three decades (Fig. 5), and there is little doubt that this

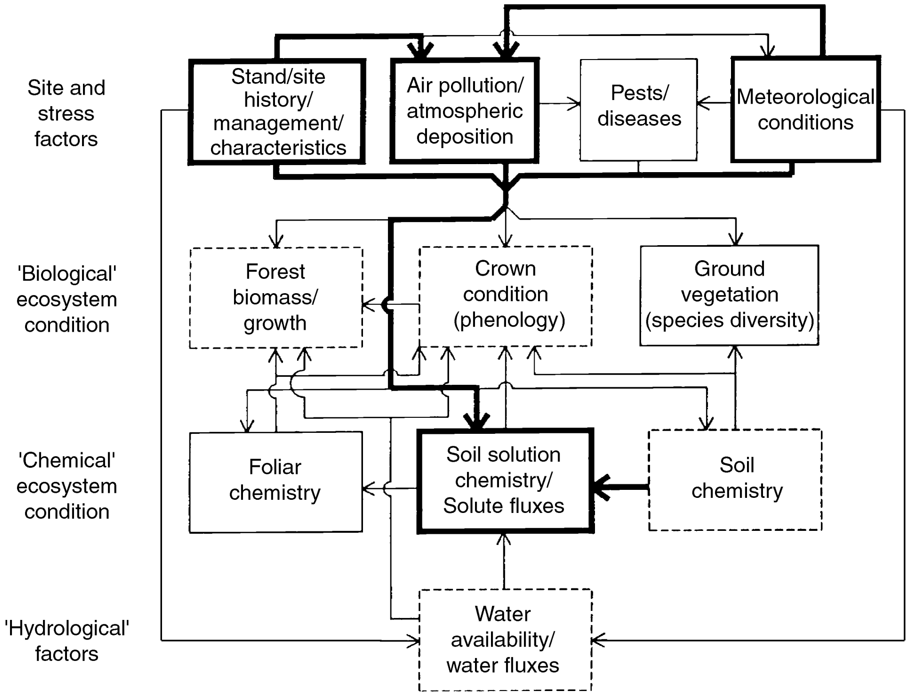


FIGURE 4 Flow diagram illustrating the relationships between site factors, air pollution, and forest ecosystem condition. From Calter de Vries *et al.*, (1988) de Vries, W., G. J. Reinds, H. D. Deelstra, J. M. Klap and E. M. Vel (1998). Intensive Monitoring of Forest Ecosystems in Europe. Technical report 1998. UN/ECE, EC, Forest Intensive Monitoring Coordinating Institute, 193 pp. with permission.

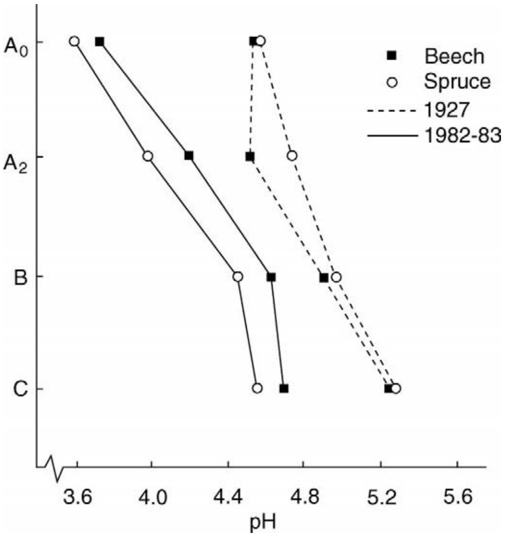


FIGURE 5 The decrease in soil pH in different soil horizons between 1927 and 1982–1983 in beech and spruce forests in southern Sweden [redrawn with permission from Hallbacken, L., and Tamm, C. O. (1985). *Scand. J. Forest Res.* 1, 219–232].

acidification is due to atmospheric deposition rather than the effects of forest growth. In several of these areas, there is evidence of adverse effects on the canopy density of major forest species such as beech and Norway spruce; these are often associated with mineral nutrient deficiencies, as demonstrated by experiments in which symptoms have been temporarily reversed by soil liming and fertilization.

However, the picture is complicated by evidence that forest growth rates are actually increasing at many European sites. Although increasing atmospheric CO_2 concentrations may contribute to this effect, there is considerable evidence that increasing rates of nitrogen deposition are a key factor, especially because growth at many of these sites is nutrient limited. The longer-term sustainability of these increased growth rates is uncertain. Experimental manipulations of nitrogen inputs to forest ecosystems at a series of sites throughout Europe during a period of 6 years have shown consistent effects on nitrate leaching, but little effect on foliar chemistry or root/mycorrhizal vitality. In contrast, field studies in Switzerland during a 10-year period have associated increased nitrogen deposition with increasing foliar nitrogen concentrations in beech and spruce foliage. There is evidence that these changes may be associated with an increased susceptibility to attacks by fungal diseases and insect pests. Growth stimulation due to increased nitrogen deposition can increase the demand for mineral nutrients such as magnesium and where these nutrients are present in limited supply can lead to foliar nutrient imbalances. Hence, the impacts of increased regional sulfur and nitrogen deposition will depend both on local site factors and on the timescale over which these effects are considered.

Most of the discussion of the consequences of soil acidification and eutrophication associated with increased sulfur and nitrogen deposition has focused on the implications for forest growth and vitality. However, this deposition may also have important consequences for woodland ground flora composition. For example, there have been reports from several areas in Europe of changes in species composition in the ground flora of deciduous forests during the past two or three decades. As expected, some basophilic or neutrophilic species have decreased in frequency or been lost, whereas some acidophilic and nitrophilic species have increased in frequency.

IV. AIR QUALITY GUIDELINES TO PROTECT BIODIVERSITY

Air pollution does not respect boundaries, and the long-range transport of air pollutants means that measures to protect biodiversity in specific areas, such as national parks or nature reserves, may have little value. Thus, many of the national parks in the United States are affected by ozone pollution, and many sites of high conservation in upland areas of the United Kingdom have been affected by acidification. It is clear that the prevention of adverse effects on biodiversity requires measures to reduce emissions and to manage air quality in the regions of concern.

However, complete control of atmospheric emissions is impossible, and the costs of decreasing pollution emissions typically increase greatly as the degree of removal increases. Thus, the following question often arises: How far should we go in reducing pollutant emissions in order to protect biodiversity? Ideally, it would be possible to define air quality standards for biodiversity, as for human health, and devise cost-effective emission control programs to ensure that these standards are met. In Europe, for example, the concept of a critical load has been introduced and is defined as “a quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge.” The critical load refers to the pollutant input in all forms of deposition; however, analogously, critical levels, defined in terms of atmospheric concentrations over a given averaging time, have also been defined for pollutant gases. It is important to appreciate that the critical load concept is based on a precautionary approach and does not relate to current levels of impact; the concern rather relates to the longer term cumulative effects of deposition. Thus, where current deposition rates exceed critical loads, action to reduce emission rates may still be justified to prevent future damage, even though field studies demonstrate no adverse effects on biodiversity of current deposition rates.

Although the critical load concept has proven to be of value in providing an effects-based approach to international negotiations on transboundary air pollution in Europe, it is important to appreciate the difficulties in defining thresholds, in terms of pollutant deposition or concentration, for adverse effects on biodiversity. These difficulties arise from three major sources:

1. There are conceptual problems in defining what is meant by a "significant harmful effect," both scientifically and in terms of societal judgment. For example, are small shifts in species composition of urban lichen communities, changes in species composition in the microfauna of a woodland floor which do not affect nutrient cycling, or small changes in the genetic composition of populations significant harmful effects. This issue often needs to be addressed in areas where other human impacts, for example, through changes in land use, have already changed substantially the structure and composition of plant and animal communities.
2. Although the adverse effects of air pollution may be clear in situations in which they are present in high concentrations and have dramatic effects, as a "threshold" concentration is approached the effects become gradually more subtle and are difficult to detect in field observations. Furthermore, in many cases the threshold concentration may be close to the natural background concentration of the pollutant or there may be no obvious threshold. Experimental methods may be adequate for assessing direct effects on single organisms, but the long-term effects of air pollution on complex communities are beyond the scope of current experimental approaches.
3. Effects on biodiversity of chronic exposures to air pollution may be the result of the long-term accumulation of pollutants or long-term chemical changes in soils, vegetation, or waters. There is no obvious experimental method of directly testing the consequences of these cumulative effects for biodiversity. The alternative approach, which has been adopted to define critical loads to prevent long-term damage to ecosystems, is to define a critical chemical concentration for biological effects in the relevant medium (typically soil or fresh water) and then to develop mathematical models to estimate the level of atmospheric deposition which would eventually lead to this chemical criterion being exceeded.

V. CONCLUDING REMARKS

This chapter summarized some key aspects of air pollutant impacts on biodiversity. Although the text has, for the sake of clarity, focused on a limited range of pollutants, many of the same principles and ideas apply in the case of other air pollutants and other ecological situations. Although in the past many situations occurred in which high concentrations of pollutants were the dominant factor causing local changes in biodiversity, most of the situations of concern today tend to involve more widely dispersed pollutants at lower concentrations, the effects of which may only become apparent over many years or even decades. In such cases, the pollutant is no longer the single dominant factor but one of a range of biological, climatic, and edaphic factors which may influence biodiversity. If we are to fully understand the role of air pollution in such situations, it is essential that we gain further understanding of the ways in which pollution can interact with these other factors and base our analysis of pollutant impacts more clearly within an ecological framework. It is also vital that we develop more effective methods to identify those areas of the planet in which current or future emissions of air pollutants are a threat to biodiversity.

See Also the Following Articles

ADAPTATION • ATMOSPHERIC GASES • GREENHOUSE EFFECT • NITROGEN AND NITROGEN CYCLE • POLLUTION, OVERVIEW

Bibliography

Aerts, R., and Heil, G. W. (Eds.) (1993). *Heathlands: Patterns and Processes in a Changing Environment*. Kluwer Academic, Dordrecht.

Barker, J. R., and Tingey, D. T. (Eds.) (1992). *Air Pollution Effects on Biodiversity*. Van Nostrand Reinhold, New York.

Bates, J. W., and Farmer, A. M. (Eds.) (1992). *Bryophytes and Lichens in a Changing Environment*. Oxford Science, Oxford.

Bell, J. N. B., McNeill, S., Houlden, G., Brown, V. C., and Mansfield, P. J. (1993). Atmospheric change: Effect on plant pests and diseases. *Parasitology* **106**, 811–824.

Freedman, B. (1995). *Environmental Ecology*, 2nd ed. Academic Press, London.

Grennfelt, P., Rodhe, H., Thornelof, E., and Wisnieski, J. (Eds.) (1996). *Acid Reign '95*, Vol. 1. Kluwer Academic, Dordrecht.

Sanderman, H., Wellburn, A. R., and Heath, R. L. (Eds.) (1996). *Forest Decline and Ozone*. Springer, Berlin.

Taylor, G. E., Jr., Pitelka, L. F., and Clegg, M. T. (Eds.) (1991). *Ecological Genetics and Air Pollution*. Springer-Verlag, New York.

Wellburn, A. (1994). *Air Pollution and Climate Change: The Biological Impact*, 2nd. ed. Longman, Harlow, UK.

ALPINE ECOSYSTEMS

Christian Körner
University of Basel

GLOSSARY

alpine Refers to the life zone above the climatic high-elevation treeline, irrespective of latitude. Though originating in the Alps (a pre-Roman word for high mountains), the term is applied globally. The reader should be aware that this term is often used in a much wider sense in common language, and is also applied to regions with mountains in general, including settlements and resorts, which is not the meaning here.

apomixis A very common but "hidden" mode of clonal propagation by seeds, the embryos of which are 100% genetic copies of the source plant. Seeds are produced without fertilization, but often pollination is required to induce apomictic seed production. Apomictic plants also reproduce sexually, but these are very rare events.

clonal growth A vegetative mode of propagation and expansion of plants by runners, tillers, or plant fragment dispersal, which is very important in the alpine life zone. (See also apomixis.) Clonal plants also produce sexual offspring by seed, but their clonal propagules often show higher survival. All clonal offspring of one source plant have the same genome and hence belong to the same "genet."

ecotype or ectotypic Refers to genetic (evolutionary) differentiation within a given species (a specific "race") that reflects an obvious advantage in a given environment. Ecotypic traits are retained when individuals are transplanted into a different environment where these traits have no advantage.

life-form The size and stature of a plant under natural life conditions. Environmental constraints can cause the life-form to differ substantially from that of a plant that develops under more favorable conditions, where genotype morphology, the "growth form," finds full expression. For instance, the growth form "tree" may be modified to the life-form "shrub" in the alpine zone.

microclimate The climate that plants, small animals, and soil microbes experience, and that differs substantially from the "macroclimate" reported by weather stations. This difference is related to surface warming by the sun or cooling at night, as well as wind shelter effects, and is largely driven by relief, exposure, ground cover, and plant stature.

treeline Also known as the forest line, this describes the high-elevation limit of (usually fragmented) forest. Most often there is no "line" visible in the landscape, so treeline position represents a convention. "Outpost" tree individuals may occur at higher elevations (the tree species line), and the boundary of closed, tall forest with timber-size stems (the timberline) commonly occurs at an elevation that is 50–150 m lower. The whole transition from timberline to the tree species line is called the "treeline-ecotone."

BIOLOGICAL DIVERSITY IN THE ALPINE LIFE ZONE encompasses all life-forms that exist above the climatic high-elevation treeline. The emphasis in this article is on plants, but microbial and animal diversity are briefly commented on as well. The alpine "biome" is unique in occurring worldwide at all latitudes, though at different elevations. The areal extent of this biome is discussed first, followed by consideration of its environmental parameters and the major drivers of its biological richness. Alpine biodiversity will then be considered from a morphological, physiological, and taxonomic point of view. A discussion of the diversity of structures and functions will lead to a brief account of the ecological function of alpine biodiversity. Finally, this article will summarize the potential threats to alpine biodiversity from global environmental change.

I. THE ALPINE LIFE ZONE

The lower boundary of the alpine life zone is, by definition, the natural climatic high-elevation treeline. Where a treeline is missing, as is the case in some dry continental areas or because of deforestation, the elevation of the nearest existing natural treeline is taken as a guideline. This is nothing more than a practical convention. The treeline and hence the lower end of the alpine zone do not form a sharp boundary, for patches of stunted trees and alpine plants often intermingle. The common climatological denominator of this boundary is a mean temperature during the growing season of 5–7°C; this temperature applies to alpine habitats worldwide (see the discussion in Körner, 1999). The length of the growing season varies from 12 months in the tropical alpine zone to merely 4–6 weeks in alpine snowbed vegetation at higher latitudes.

Between 5 and 6 million km^2 fall into the alpine life zone (ca. 5% of the vegetation-covered land area of the globe). Because not all of this alpine area is covered by vegetation—some consists of bare rock, rock fields, scree, and glaciers—the vegetated alpine land area is estimated to be 3 to 4 million km^2 (see Körner, 1995).

Two-thirds of the global alpine area is situated in the temperate and subtropical zones and only 10% occurs in the tropics (Table I). Partly this is because mountains in the tropics need to be very high, at least 3600 m, to permit tropical alpine vegetation to occur, whereas at the polar circles mountain heights of only 600 m are required to support alpine habitat. These are the approximate treeline elevations of the respective climatic zones. At about 4000 m, treeline reaches its highest elevations in the subtropics (Fig. 1). At high latitudes (>65 to 70°N), alpine vegetation merges with arctic tundra. Despite a number of common taxa and the overwhelming influence of low mean temperature, the arctic tundra life zone is very different from the alpine zone in terms of climate, land surface structure, and vegetation, hence several authors (e.g., D. Löve and W. D. Billings) have recommended that alpine vegetation not be referred to as "alpine tundra." The upper limits of vascular plant occurrence are commonly 1000 to 1500 m above the lower limits of the alpine zone (the treeline), but some extreme high-elevation outposts of higher plants are found up to 4400 m in the temperate zone and up to 6200 m in the subtropics, where the uppermost individual of a higher plant was found in the Himalayas (Miehe, 1997). An important feature of alpine life is its isolation. Mountaintops with their high biological diversity represent islands or archipelagos surrounded by lowlands, where most alpine organisms cannot survive.

Various aspects of plant diversity in high mountain systems have been reviewed previously. A selection of such synthetic papers or volumes is included in the Bibliography. Full bibliographic references to original studies (mentioned in the text by authors' names only) can be found in Körner (1999).

TABLE I

The Global Distribution of the Alpine Life Zone

Climatic zone	Latitudinal range Northern hemisphere	 Southern hemisphere	Percent of total area
Subarctic/subantarctic alpine	>60°N	>50°S	23%
Temperate alpine	40–60°N	35–50°S	33%
Subtropical-meridional alpine	25–40°N	20–35°S	34%
Tropical alpine	0–25°N	0–20°S	10%

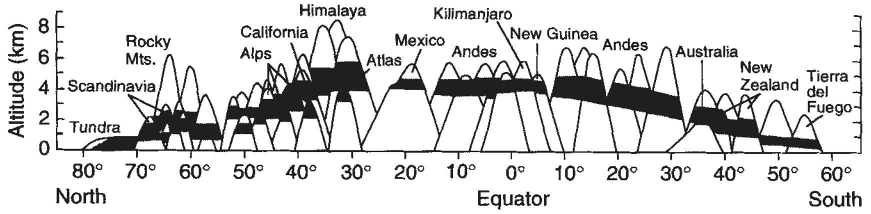


FIGURE 1 The alpine life zone occurs at all latitudes, though at contrasting elevations. (From Körner, 1999.)

II. HABITAT DIVERSITY, A KEY TO ALPINE PLANT DIVERSITY

A. Climatic Microhabitats

There is a common belief that the alpine life zone is very hostile to plants and animals and that low temperatures restrict life activities, including productivity. This generalization is wrong for two reasons: (1) as will be shown, temperatures are not always that low, and (2) this belief reflects our human perspective of what is "cold." For organisms adapted to alpine life conditions the temperatures they experience are not necessarily "cold." In fact, if temperatures were higher, most of these organisms would suffer or even die, many of them because they would be outcompeted by other species that do better at such higher temperatures.

The most important characteristic of the alpine life zone is its fragmentation into a multitude of microhabitats created by relief, exposure, and slope, which interact with solar radiation and wind to cause soil moisture, temperature, and substrate quality to vary enormously over very short distances. It is this diversity of microhabitats, the steepness of the terrain in particular, that makes the alpine life zone so different from arctic tundra, and that is responsible for its much greater organismic richness. Across a few meters one can easily find snowbed plants, just emerging from melting snow, and succulent plants on outcrops, which perform crassulacean acid metabolism (CAM) as do many hot desert plants. In fact, these succulents may regularly experience temperatures close to 50°C on steep south slopes under full midday sun, even at high latitudes.

In addition to these microclimatic determinants imposed by land surface structure, plants themselves influence their micro-environment (Fig. 2). Depending on life-form, plants may decouple themselves from ambient air conditions. Under direct insolation, compact cushion plants or prostrate dwarf shrubs have been shown to warm up to tropical temperatures. These life-form-dependent microclimates largely disappear under thick clouds or at night, but storage of warmth in the topsoil is also strongly influenced by the type of plant cover, and so is radiative cooling during clear nights.

Hence, the climate that alpine plants experience can be very different from what might be expected based on elevation alone or data measured at weather stations. Such weather data monitor the temperature, humidity, and windspeed that occur outside the calm boundary layer into which alpine plants (leaves in particular) and their animal and microbe partners are commonly nested. A widespread assumption is that alpine plants are small and prostrate because their growth is temperature limited. The reality is that, because alpine plants are small (and mostly stay small when grown at higher temperatures), they may periodically escape the cold and, at least during sunny hours, experience radiative warming, and thus are not always colder than lowland plants. This is well reflected in their thermal optima for photosynthesis, which (in the temperate zone) were

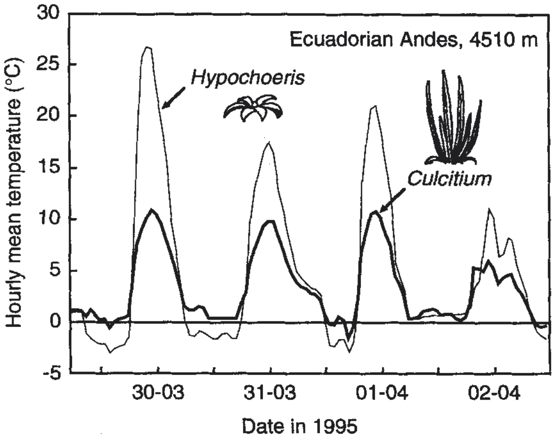


FIGURE 2 The diversity of plant morphologies creates diverse microclimates. Note the different temperatures in upright and prostrate leaves of two species in the Ecuadorian Andes. (After M. D. Diemer, in Körner, 1999.)

shown to be similar to those of lowland plants. In contrast, night-time temperatures may be prohibitive for plant growth at alpine elevations, and thus limit structural investments of assimilates acquired during the day. Because of the strong link between plant life-form and microclimate, alpine biodiversity can be understood only if one accounts for plant structural diversity.

B. Diversity of Substrates

Alpine microhabitats may belong to a suite of different land surface structures and soils, the ten most important of which are common to all mountains:

- exposed rock terraces and rock crevices
- block fields
- scree and mixed scree/rock slopes or flats
- drained ridges and plateaus
- periodically wet depressions, gullies, or snowbeds
- gentle slopes with relatively stable soil
- steep slopes with creeping soil
- flats with cryogenic structure (e.g., hummocks, polygons)
- mires or other wet ground
- springs, water flows, or lakes

These ten habitat types are not exhaustive. Special habitats not found everywhere include sand drifts or dunes and salt flats (in some semiarid subtropical mountains), disturbed surfaces due to animal trampling, rest places, or burrowing, avalanche tracks, and man-made landscapes such as pastures.

Each of these types of habitats may be found at different exposures to sun and wind (so the number of substrate types multiplies with the number of microclimatic conditions, despite some redundancy), and each of these combinations in turn may be found on very different parent rock material such as calcareous or siliceous, mixed metamorphic, or volcanic material. Depending on wetness and elevation, plants may have converted the top layers of the substrate to an extent that it becomes chemically independent from the parent rock (e.g., humic turf of pH 3.7 overlaying calcareous bedrock). Varying degrees of erosion and soil formation will then enhance the spatial heterogeneity of the substrate. The availability of soil nutrients and the extent of soil development are strongly relief driven, but plant life-forms also determine local nutrient retention. By forming dense cushions or tough tussocks, plant litter and thus the precious nutrients contained in it are prevented from being blown or washed away. When trapped beneath or strongly attached to the plant, litter can decompose and nutrients can recycle locally. Plant life-form, root and rhizome structure in particular, also influences substrate stability and soil formation (see Section V).

To understand alpine biodiversity, it is important to be aware of this extensive microenvironmental patchiness of the alpine life zone, and the capacity of plants to influence their life conditions substantially. Taken together, the foregoing combinations of microenvironmental conditions yield hundreds of very specific niches, each preferred by a different combination of species.

III. PLANT DIVERSITY IN THE ALPINE LIFE ZONE

A. Diversity of Morpho-types

Though the alpine life zone is not as rich in life-forms as a humid tropical forest, the diversity of morphotypes found here is surprisingly high. There are ten principal groups of life-forms, eight of higher plants and two of cryptogams, irrespective of whether individuals perform clonal growth (see later). The first four groups are most important:

- low stature or prostrate woody shrubs;
- graminoids such as grasses and sedges, many forming tussocks;
- herbaceous perennials, often forming rosettes; and
- cushion plants of various types.

Less common or of more regional importance are:

- giant rosettes of tropical mountains;
- geophytes, mainly confined to mountains with a pronounced seasonality;
- succulents, with both stem and leaf succulence; and
- annuals (sometimes biannuals), which become quite rare at high elevations.

The remaining two life-forms are cryptogams, that is, desiccation-tolerant, non-flowering plants:

- bryophytes ("mosses"), in some areas also ferns and lycopods; and
- lichens (including fruticose, foliose, and crustaceous).

These life-forms, in mixtures of varying abundance of each, compose the "alpine vegetation." In addition, algae and fungi play an important role. The diversity of plant structures is further enhanced by various modes of clonal propagation, which become increasingly important as elevation increases. The following is a short list of the diversity of clonal structures (Fig. 3):

- tussock graminoids
- stoloniferous graminoids
- mat- or cushion-forming forbs
- stoloniferous or rhizomatous forbs
- creeping dwarf shrubs
- prostrate dwarf shrubs
- viviparous plants
- accidential clonal plants (fragmentation by external forces)

These typologies do not account for plant height or degree of horizontal spreading, both of which vary considerably among species and microclimates. Overall, the diversity of plant stature in the alpine life zone is perhaps nearly as large as the taxonomic diversity (see

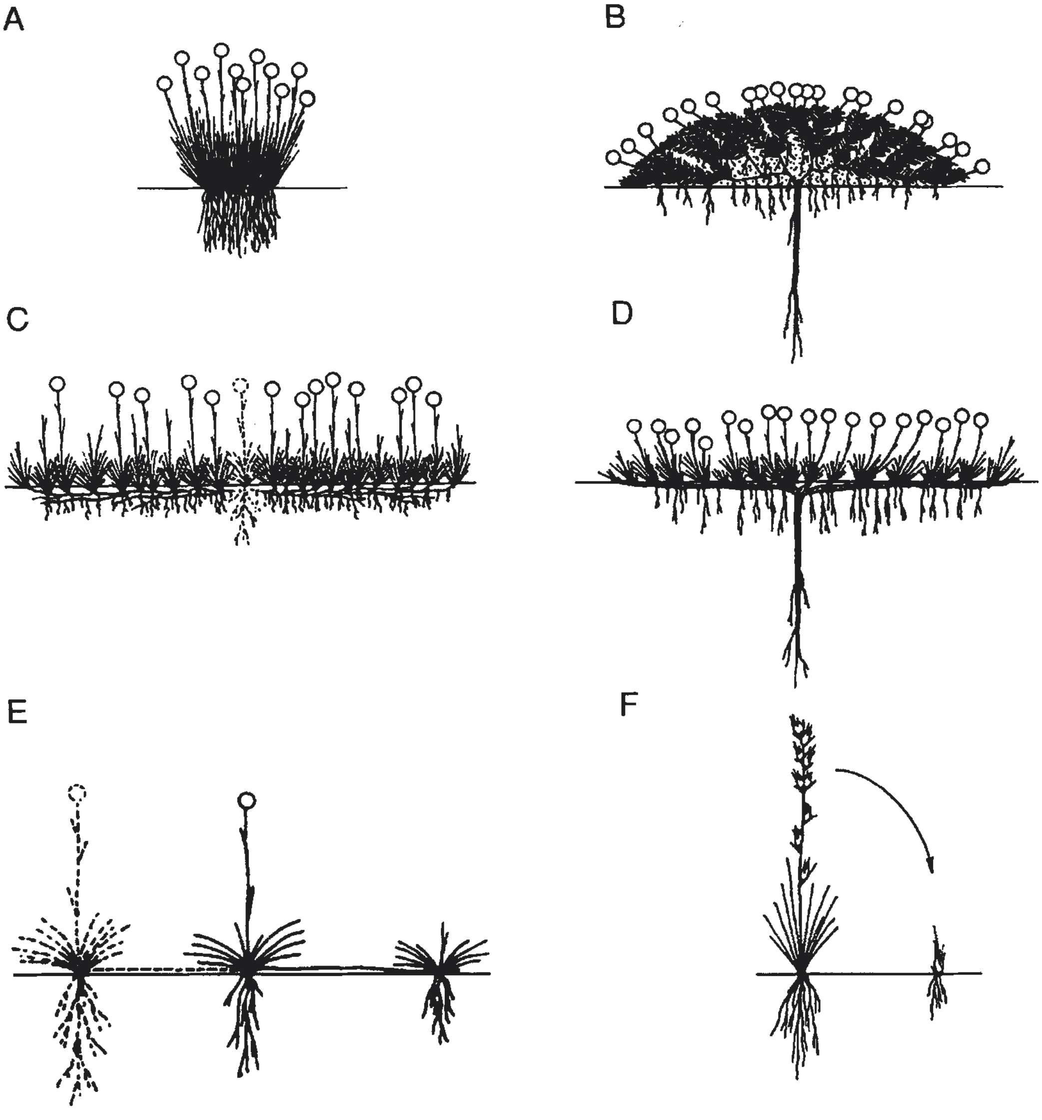


FIGURE 3 Diversity of clonal growth in alpine plants. The examples shown here include (A) graminoid tussocks, (B) cushions with adventitious roots and potential later fragmentation, (C) mats of rhizomatous forbs or graminoids, (D) mats of proliferating forbs that retain a primary root, (E) stoloniferous plants, and (F) "viviparous" plants producing bulbils or floral plantlets. (After H. Hartmann, from Körner, 1999.)

Halloy and Mark, 1996). It is obvious that these structural features are significant functional attributes, strongly associated with microhabitat preference and/ or microhabitat tolerance.

B. Diversity of Physio-types

The term physio-type is used here to circumscribe the physiological attributes of alpine plants, which may be as diverse as the variety of life-forms. Certain morphological attributes are functionally linked to physio-types. The following is only a brief summary. Readers with an interest in this field will find an extended treatment of the subject in Körner (1999).

Rates of plant photosynthesis, how photo-assimilates are invested in the plant body, nutrient use, water relations, stress resistance, and secondary metabolites all vary substantially among alpine species. The predominance of physical limitations to growth, particularly at the uppermost limits where plants can grow and survive, may be expected to narrow the spectrum of possibilities, in the ultimate case by permitting only one way to survive. Surprisingly this is not so. Even in habitats that by all standards can be rated as "extreme," one can find a suite of physio-types (often associated with specific morpho-types). This is a most important point for the study of alpine plant adaptation. The selection of a single species for study inevitably will produce data with no generalization potential. "*The* alpine physio-type" does not exist. One randomly selected species or small group of species may represent a very special case, and be all but "typical."

A good example is the way plants invest in biomass. Although this finds expression in morphology, the quantitative aspects of it are directly related to plant metabolism. Plants may favor roots, stems, storage organs (all three are net sinks for carbon), or leaves (the net source of carbon). How plants invest is key to the understanding of whole-plant carbon balance and to growth and reproduction. Co-occurring species, equally successful in terms of abundance and often found together in the same habitat, may represent the left and right tails of a frequency distribution of these traits at a common attitude (Fig. 4). There seems to be a multitude of ways to cope with the demands of life conditions even at extremely high elevations.

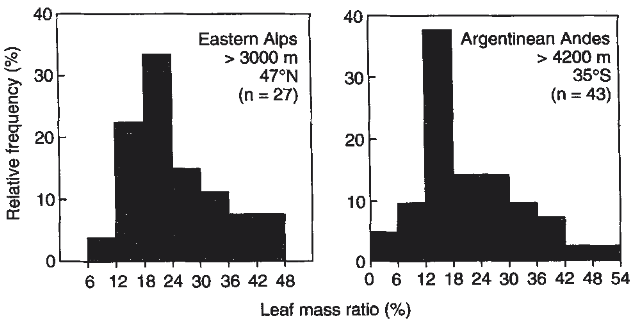


FIGURE 4 Frequency distribution of leaf mass ratio (LMR, the fraction of leaf dry matter versus total plant dry matter) in the Alps and the Andes. In the Alps, for instance, co-occurring species are found in the left (*Ranunculus glacialis*) and right (*Cerastium uniflorum*) tails, reflecting the great diversity of plant dry matter investment even at extremely high elevations. The same is true for any other mountain region, including the example for Argentina shown in the right diagram. (Modified from Körner, 1999.)

Another example is the way alpine plants construct their leaves in terms of leaf dry matter investment per unit leaf area. "Expensive" (thick) leaves contain a lot of dry matter per unit leaf area, whereas "cheap" (thin) leaves contain little. Again the diversity of this trait expressed as "leaf mass per area" (LMA, or its reciprocal specific leaf area, SLA) is very large. There is nothing like a typical alpine SLA (Fig. 5; for a full account of such leaf traits in a global comparison see Körner *et al.*, 1989). The same applies to leaf nitrogen concentration and mobile carbohydrates, although there are slight overall trends with increasing elevation for SLA to decrease and nitrogen and mobile carbohydrate concentrations to increase (leaves becoming more "expensive"). However, such means across many species need to be treated with care, given the great diversity. It cannot be concluded that alpine plants have lower SLA when a substantial fraction of all studied alpine species does not fit this pattern. The species from different elevations that are included or excluded from such community subsamples will always affect the mean.

The same applies to any other physiological trait that has been studied in more than one species. For example,

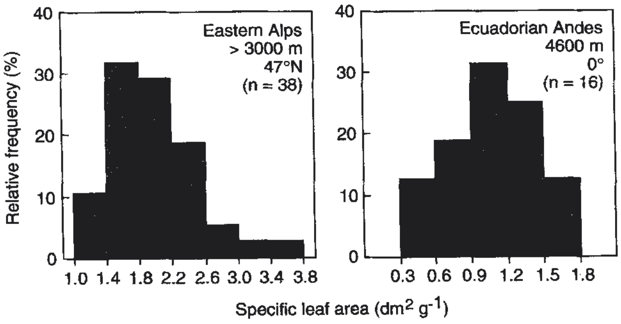


FIGURE 5 Frequency distribution of specific leaf area (SLA) in alpine species in the Alps and the Andes. (From Körner *et al.*, 1989.)

freezing resistance varies enormously in alpine plant species. According to W. Larcher, some species can survive any low winter temperature that could occur on Earth (less than −70°C), and others are killed by only −12°C. Freezing resistance is one of the few traits where this diversity of responses could be explained by habitat characteristics, in this case the predictability of snow cover in winter. But for most other traits such clear-cut causal links have not been found. It rather seems there are many different ways to cope with similar problems. These different traits may be "non-functional" in the sense of not being critical for survival and reproduction under "normal" situations, but some may become decisive under very specific and rare conditions through which they might have been selected for as advantageous (see the discussion of the ecological function of diversity in Section V).

C. Diversity of Reproduction

Recruitment via seed is rather risky in alpine environments because the seedling needs to establish in an often short season and on a potentially hostile substrate. Hence the most common strategies for alpine plants to survive and persist are long life and clonal propagation. The production of seed itself does not seem to be a problem and alpine plants produce lots of viable seed in most years, as has been known from the beginning of alpine plant research in the nineteenth century (see review by Körner, 1999). For a functional understanding of alpine biodiversity, seed production is very important, as rare as successful seedling establishment might be. It is through sexual reproduction that the genetic diversity is retained and by which the ecotypic differentiation of traits, essential in the fragmented and ever-changing alpine environment, develops. The most important pollinators are flies; at lower alpine elevations, bumblebees, solitary bees, and butterflies become more prominent (with decreasing importance by order; a detailed analysis for the southern Andes was done by M. K. Arroyo). Roughly one-fourth of alpine plant species are wind pollinated (such as Poaceae, Cyperaceae, and Polygonaceae).

Current evidence suggests that outbreeding is the dominant form of sexual reproduction, but most alpine plants can also self (which secures some seed production if pollen transfer fails because of bad weather), and some are even obligatory selfers. The most extreme forms of retaining the parental genome is apomixis, where embryos are genetic copies of the source plant. Surprisingly, both clones and apomicts exhibit high intraspecific genetic diversity, indicating that sexual reproduction also occurs in such plants. Through genetic fingerprinting, T. Steinger and coworkers recently showed that some of these obligatory clonal alpine plants can produce very large genets and be thousands of years old, but clones of different genets were found to be intermingled. Yet clonal propagation must not be seen as a substitute to sexual propagation. Both the sexual and clonal modes occur simultaneously for most of the time. There is no indication of genetic depauperation within species with increasing elevation, but this is a field that needs more research. Current knowledge does not suggest that alpine plant diversity is limited by reproductive constraints.

It is well established through classic transplantation experiments in California, the Rocky Mountains, and the Alps that ecotypic differentiation does occur among populations from different elevations, so that genotypes of the same species from high elevation differ in stature, phenology, and physiology from lower-elevation genotypes when grown in a common garden (see review by Clements *et al.*, 1950). Hence, in addition to species diversity, discussed next, there is also a high degree of within-species diversity.

D. Taxonomic Diversity

The total alpine flora of the globe consists of approximately 8000 to 10,000 species of flowering plants (Körner, 1995). These belong to about 100 (±10) families and 2000 genera. Hence, one-fourth to one-third of all plant families of higher plants have representatives in the alpine life zone. Assuming a total known global flora of 250,000 species, alpine species contribute about 4% of the global plant species diversity. Given that only 3–4% of the global land area that is suitable for at least some plant growth falls in the alpine zone, global alpine plant diversity per unit land area corresponds to the global mean of other biomes. Although such gross means need to be treated with great caution, one can at least conclude that plant species diversity in the alpine life zone is comparatively high, in view of the generally less luxurious life conditions found there.

Whether high mountains are biodiversity hot spots (Barthlott *et al.*, 1996) is a question of scale. The preceding numbers strictly refer to the alpine life zone as defined in Section I. If one selects a census area that includes the full spectrum of elevations (e.g., a cross section through a whole mountain range), it may encompass almost all life zones on Earth from humid tropical forest at the bottom to glacier forefields at the top. High mountains, those in tropical latitudes in particular, indeed represent an incredible compression of biomes. Climatic and vegetation zones separated by several thousands of kilometers at sea level may be found within 50 km or less horizontal distance on the flanks of the major subtropical or tropical mountain systems. However, the topic of life zone compression in mountains exceeds the frame-work of this article, which is restricted to the discrete alpine part of mountain vegetation.

A single mountain system such as the Rocky Mountains, the Alps, the Caucasus, the Venezuelan part of the Andes, or the mountains of New Zealand commonly has an alpine flora consisting of 600 to 1500 species. This needs to be considered in view of a total flora of the arctic tundra of about 1000 to 1500 species (depending on how sub-species are ranked). In this respect, biodiversity of the alpine life zone is outstanding. A distinct mountain region such as the Teton Range in Wyoming, the Snowy Mountains of Australia, or the central part of the Swiss Alps commonly contains roughly 200–400 alpine species, a number that is surprisingly constant across the globe (most ranges harbor around 250 species). On a single sample plot (e.g., 100 × 100 m) one may find one-third of the total regional flora. An analysis of the Swiss Alps by T. Wohlgemut revealed that alpine plant species diversity increases with the size of the observation area up to 20 km^2, but beyond that species numbers level off.

Plant species diversity generally decreases with increasing elevation (Fig. 6a), although there may be intermittent peaks where two altitudinal life zones merge. Again this is a question of scale. The decline of species diversity is particularly impressive in the uppermost part of the alpine life zone (Fig. 7).

The diversity of cryptogams, mosses and lichens in particular, is relatively high in the alpine life zone. Species numbers may be of the same order of magnitude as for vascular plants, depending on humidity. With greater humidity one generally finds higher numbers of cryptogams. In some parts of temperate zone and subarctic mountains the biomass of fruticose lichens in alpine grassland is double that of higher plants. Since mosses and lichens are desiccation tolerant, they can occupy bare rock and scree and are among the first organisms to initiate humus accumulation on raw substrates. Although cryptogams have almost no elevation limits as long as there is some bare substrate (rock lichens are reported at 7200 m elevation in the Himalayas), the number of cryptogam species also declines with greater elevation, as is shown in Fig. 6b for bryophytes.

Besides the major role of microhabitat diversity discussed earlier, there are also historical reasons for alpine plant diversity. While the lowland flora changed owing to global climatic changes, such as after the ice ages, a fraction of the then cold-adapted lowland flora migrated to high elevations and enriched the existing older stock of alpine species. This is how the edelweiss (*Leontopodium alpinum*) became a recent addition to the alpine flora of the Alps when the ice retreated. Agakhanjanz and Breckle (1995) suggested that in some mountains (e.g., the Pamirs) tectonic uplift matched time constants of speciation and contributed to the autochthonous species richness.

In summary, the alpine life zone is rich in plant species that belong to a multitude of morphological, physiological, and reproductive types and that inhabit a broad spectrum of microhabitats, reflecting both his-

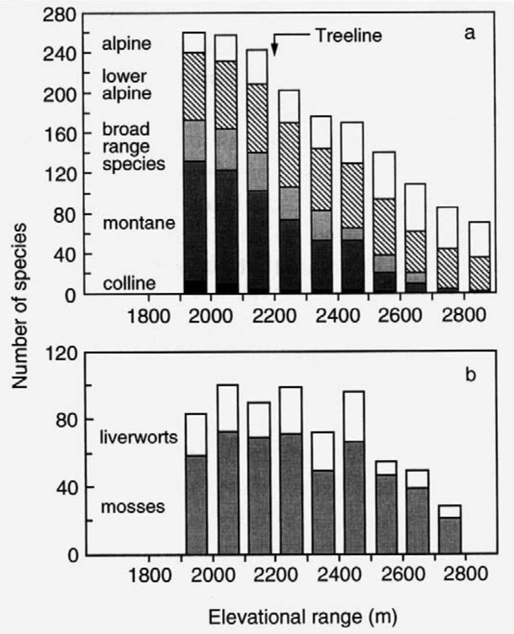
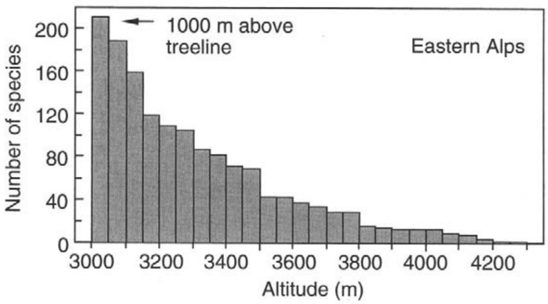


FIGURE 6 Altitudinal trends in species diversity of higher plants in the Alps (a, after J.-P. Theurillant and A. Schlüssel) and bryophytes (b, after P. Geissler and C. Velluti). (Adapted from Körner, 1999.)



FI [illegible] ermost range of higher plant life in the Alps. (After G. Grabherr, from Körner, 1999.)

torical and even geological events. Because alpine plants are small, species diversity may exceed 50 species per 1 m^2 of land area, which is among the highest in the world. Alpine plants also provide a varied diet of food for alpine animals and microbes, the diversity of which will be briefly touched upon in the following section.

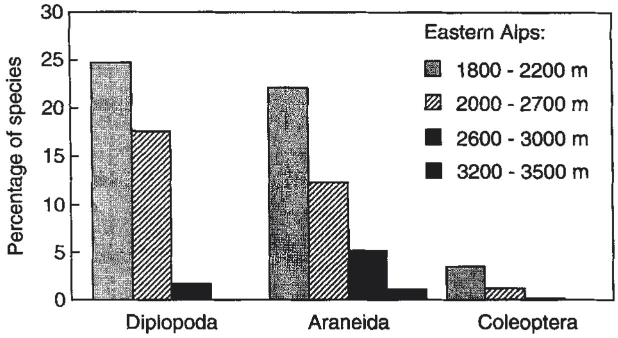


FIGURE 8 Examples of elevational trends in invertebrate species diversity in the Alps. (From E. Meyer and K. Thaler, 1995.)

IV. DIVERSITY OF ALPINE ANIMALS AND MICROORGANISMS

A. Animal Diversity

The great variety of animals from mites to birds, and the greater fragmentation of expertise in animal sciences, may explain why there has been no attempt at a global synthesis of animal diversity in the alpine life zone. Franz (1979), in his German book on high mountain ecology, referred to a great number of otherwise scattered observations, and Meyer and Thaler (1995) attempted a summary for invertebrates, with a focus on the Alps. The following brief account leans heavily on the latter publication.

The only resident vertebrates in the uppermost part of the alpine zone (often called the nival zone) are voles, with the record finding by S. R. P. Halloy at ca. 6000 m elevation on Sucompa volcano in the Argentinean Andes. Snow voles are very abundant and active even above 3200 m in the Alps. At these elevations birds are visitors, but only a few hundred meters lower they are resident as the plant cover becomes more regular. In the Argentinean Andes, ducks breed at 4250 m elevation (personal observation at an alpine lake in the Cumbres Calchaquies). Snakes, lizards, and frogs are reported to occur up to 1000 m above treeline. Large mammal grazers (e.g., ibex in the Alps and guanacos in the southern Andes) occur at almost any alpine elevation, except for peaks surrounded by glaciers or inaccessible cliffs. These grazers profoundly influence the development of alpine plant diversity. In the lower and middle alpine belts, wild herbivores have been replaced by domestic herbivores in many areas (sheep, goats, yaks), which continue to exert selective pressure on vegetation. Some of these animals graze the highest fragments of vegetation, as was reported for yaks, which were found grazing above 5000 m in the dry part of the Himalayas (Miehe, 1997).

Quantitative data for some important groups of invertebrates have been compiled by Meyer and Thaler (1995). The major losses of taxonomic groups as elevation approaches the upper limits of closed vegetation are those of earthworms, gastropods, grasshoppers, Hymenoptera, and beetles (Fig. 8), with the latter showing the greatest decline already below treeline. By contrast, flies, spiders, and springtails remain quite abundant (at 300 m above treeline, one-fifth of the total regional number of 200 springtail species were found in the Tirolian Alps). According to L. W. Swan, a salticid spider species was collected in the Himalayas at an altitude above any plant growth, possibly living on the aerial import of small arthropods from lower elevations. Only 2% of the 400 species of spiders of the central Alps of Tirol (Austria) are regularly found in the uppermost alpine zone. The more open vegetation becomes, the more invertebrate life is linked to the occurrence of compact plant forms, such as cushion plants.

In their overview, Meyer and Thaler noted that among invertebrates the herbivore species numbers declined more rapidly than species numbers in higher trophic levels. The reduced species number at "extreme" habitats is often balanced by greater individual densities per species. Overall, animal species diversity in the alpine life zone follows similar elevational trends as in plant species diversity, but in terms of the total numbers of alpine species, animals may exceed plants by factors of 5 to 10 (this is a personal guess). Establishing such diversity ratios would be an important contribution to the understanding of alpine biodiversity in general.

B. Microbial Diversity

As the altitude above treeline increases, animals make a smaller contribution to plant litter decomposition and soil formation and microbes become more important, even as their species numbers decline steadily. Diversity of soil fungi decreases with altitude and mycorrhization, the important plant root–fungus sym-

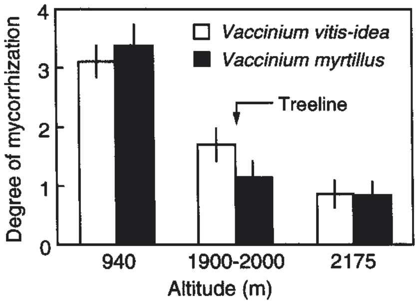


FIGURE 9 Elevational trends in mycorrhizae occurrence in the Alps. (After K. Haselwandter and D. L. Read, from Körner, 1999.)

biosis, also decreases (Fig. 9). All known types of mycorrhizae occur in alpine soils: ectomycorrhizae (e.g., on *Salix, Dryas, Polygonum,* and *Kobresia* spp.), ericoid mycorrhizae (Ericaceae), vesicular–arbuscular (VA) mycorrhizae (most forbs, grasses, and some sedges), and even orchid mycorrhizae. Non-mycorrhizal plant species can also be found (Gardes and Dahlberg 1996). So-called vesicular–arbuscular mycorrhizae and dark-septate hyphae of unknown taxonomic affiliation are found even in the highest rock and scree habitats, though at greatly reduced abundance. Completely isolated plants above 3000 m in the Alps were found to be largely free of mycorrhizae, but the dark-septate hyphal root fungi could still be seen (for references see Körner, 1999).

The most robust of all organisms, the bacteria plus some unicellular fungi, retain a high diversity and abundance in the alpine life zone. F. Schinner and coworkers isolated 130 different strains of microorganisms in alpine environments of the western and eastern Alps that can survive and multiply at 0°C. Of these, 77% were bacteria, 20% yeasts, and only 3% hyphal fungi. A very detailed analysis of bacterial diversity on Niwot Ridge, Colorado, by R. L. Mancinelli showed that *Pseudomonas* and *Bacillus* were the most abundant genera. Dinitrogen-fixing as well as nitrifying and denitrifying bacteria are abundant in alpine soils. Bacteria have no altitudinal limits as long as some organic dust and short spells with liquid water occur, and Swan reported a number of taxa isolated from substrate collected at 8400 m altitude on Mt. Everest, the environment on Earth that he thinks is most comparable with that of Mars.

Taken together, these observations indicate that mycorrhizae are a common element of alpine plant life, being more prominent in lower alpine elevations and in infertile soils and becoming rare only with isolated plants on high mountain peaks, where soils have little carbon. Rich bacterial life is found even at the highest elevations, indicating a capacity for metabolism under the most extreme conditions.

V. POTENTIAL FUNCTIONS OF ALPINE BIODIVERSITY

Among the theories that attempt to explain the functional significance of organismic diversity, the insurance theory seems to be most relevant for the alpine life zone. In simple terms it says that a species-rich, functionally partly redundant organismic "work force" ensures the functional integrity of ecosystems even if some of the organisms die out. A complete loss of species by extreme stress or through pathogens would be nearly irreversible in steep alpine terrain, because it is the presence of at least some species that prevents the soil from being washed away. The sustained functional integrity of alpine ecosystems is inevitably tied to the presence of soil, and this presence is inevitably dependent on the roots and rhizomes that hold it. Strong-rooted plants are the keystone elements for the preservation of the alpine ecosystem in steep terrain. A high diversity of species is commonly associated with a high diversity of rooting patterns, which in combination create the mechanical strength required to hold the soil.

B. Messerli and J. D. Ives estimated that 10% of the global human population depends directly, and 40% of the population indirectly, on mountain ecosystems, and thus the stability of their upper part, the alpine zone, is of critical significance to society. Supplies of drinking and irrigation water, as well as the safety of hydroelectric schemes and transport routes, depend on intact upslope soil conditions, and alpine biodiversity helps to ensure ecosystem health. Of course alpine biodiversity also provides other things. Alpine meadows and fellfields are very attractive landscapes for human recreation, and with their high biodiversity they are of prime conservation value; in many regions they represent the last undisturbed natural areas. Because the alpine life zone is represented across the globe, it is also ideally suited for global monitoring of biological responses to atmospheric change.

VI. ALPINE BIODIVERSITY AND GLOBAL CHANGE

Global change has many facets, all related to the ever-increasing use of resources and land area because of human population growth and the increasing consumption of goods. The most severe impacts on alpine ecosystems are (1) land use practices, (2) potential global warming with associated changes of snow cover and permafrost, and (3) increasing wet nitrogen deposition. Other aspects of global change, such as atmospheric CO_2 enrichment by itself, increasing ultraviolet radiation due to ozone layer depletion, and air pollutants other than nitrogen loading, seem to be of minor significance on a global scale (see review by Körner, 1999).

Land use, in particular the intensification of pasturing or the reverse, the abrupt abandonment of former, traditionally pastured alpine terrain, exerts the greatest influence. The steeper the terrain, the more critical these effects become. Overgrazing generally diminishes plant diversity and ruins the protective plant cover within a few seasons, with self-repair of the ecosystem commonly occurring at a much slower rate than soil erosion. In contrast, moderate and well-managed grazing can increase biodiversity and create short and dense swards of vegetation that are extremely robust against erosion and increase catchment water yield. Some of the biodiversity hot spots of the Alps, the Caucasus, and the Himalayas are traditional pastures in the lower alpine belt. There is no way of maintaining these systems by "let alone" grazing strategies, because herds of domestic animals tend to crowd certain areas and ignore others. Maintaining these intact alpine grasslands is very important for at least three reasons: (1) their often very impressive biological richness, (2) their continued potential as a "clean" food source, and (3) their positive influence on water yield. The abrupt abandonment of pastures leads to an unstable transition phase that may last for a century before new, well-adapted communities of species are able to return. A later reversal back to the previous biodiverse pasture-land is often impossible within a reasonable time and with affordable effort, because the biological structure and the soils of these pastures took many hundreds or thousands of years to develop, but are rapidly converted.

Climate warming will affect alpine biodiversity in subtle ways because of the great variety of intermingled microhabitats (see Section II). Although Grabherr *et al.* (1995) documented some upslope migration of species, the more important changes possibly happen among microhabitats, with new niches filled by species and other niches abandoned (Fig. 10; Gottfried *et al.*, 1998). These mosaics of life conditions represent a certain margin of safety against the loss of alpine species in a slightly warmer overall climate. However, the abundance of species will change as the abundance of their microhabitats changes. Effects on snow cover and snow duration will be more critical than temperature per se.

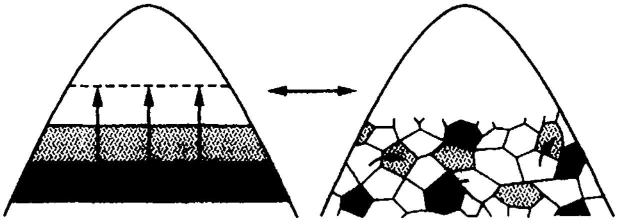


FIGURE 10 Global warming and alpine plant diversity: upslope migration of species or niche filling? Both responses are likely and will alter local species abundance. Niche filling, that is, "horizontal reallocation," will precede longer-term changes of the position of whole vegetation belts.

Exceptions to these scenarios are mountains that are not high enough, in which case the current alpine biota will find no upslope escape if it gets warmer.

Since alpine vegetation is well adjusted to cope with low soil nutrient levels, the regional increase of soluble nitrogen deposition will influence biodiversity. More vigorous and nitrogen-demanding species are likely to gain space over slower-growing, smaller species. Since these more vigorous species are commonly also less resistant to stress, nitrogen deposition can increase the sensitivity of certain ecosystems, while others (e.g., pioneer vegetation) may profit. It was shown that even minute additions of nitrogen fertilizer—less than is contained in many places in lowland rain-water—can create drastic changes in the alpine flora (Körner, 1999).

In summary, the greatest risk of loss of biological diversity in the alpine life zone is human land use. However, land use can also contribute to the maintenance of highly diverse and stable ecosystems in the lower alpine belt if sustainable management practices are applied.

See Also the Following Articles

ARCTIC ECOSYSTEMS • EUROPE, ECOSYSTEMS OF • GRAZING, EFFECTS OF

Bibliography

Agakhanjanz, O., and Breckle, S. W. (1995). Origin and evolution of the mountain flora in middle Asia and neighboring mountain regions. In *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences* (F. S. Chapin III and Ch. Körner, eds.), pp. 63–80. Springer, Heidelberg, Germany.

Barthlott, W., Lauer, W., and Placke, A. (1996). Global distribution of species diversity in vascular plants: Towards a world map of phytodiversity. *Erdkunde* 50, 317–327.

Billings, W. D. (1988). Alpine vegetation. In *North American Terrestrial Vegetation* (M. G. Barbour and W. D. Billings, eds.), pp.

392–420. Cambridge University Press, Cambridge, United Kingdom.

Bliss, L. C. (1985). Alpine. In *Physiological Ecology of North American Plant Communities* (B. F. Chabot and H. A. Mooney, eds.), pp. 41–65. Chapman and Hall, London, United Kingdom.

Chapin, F. S., and Körner, Ch. (1996). Arctic and alpine biodiversity: Its patterns, causes and ecosystem consequences. In *Functional Roles of Biodiversity: A Global Perspective* (H. A. Mooney, J. H. Cushman, E. Medina, O. E. Sala, and E. D. Schulze, eds.), pp. 7–32. John Wiley & Sons, Chichester, United Kingdom.

Clements, F. E., Martin, E. V., and Long, F. L. (1950). *Adaptation and Origin in the Plant World. The Role of Environment in Evolution.* The Chronica Botanica Company. Waltham, Massachusetts.

Franz, H. (1979). *Oekologie der Hochgebirge.* Verlag Eugen Ulmer, Stuttgart, Germany.

Gardes, M., and Dahlberg, A. (1996). Mycorrhizal diversity in arctic and alpine tundra: An open question. *New Phytol.* 133, 147–157.

Gottfried, M., Pauli, H., and Grabherr, G. (1998). Prediction of vegetation patterns at the limits of plant life: A new view of the alpine–nival ecotone. *Arctic and Alpine Res.* 30, 207–221.

Grabherr, G., Gottfried, M., Gruber, A., and Pauli, H. (1995). Patterns and current changes in alpine plant diversity. In *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences* (F. S. Chapin III and Ch. Körner, eds.), pp. 167–181. Springer, Heidelberg, Germany.

Halloy, S. R. P., and Mark, A. F. (1996). Comparative leaf morphology spectra of plant communities in New Zealand, the Andes and the European Alps. *J. Roy. Soc. New Zealand* 26, 41–78.

Körner, Ch. (1995). Alpine plant diversity: A global survey and functional interpretations. In *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences* (F. S. Chapin III and Ch. Körner, eds.), pp. 45–62. Springer, Heidelberg, Germany.

Körner, Ch. (1999). *Alpine Plant Life.* Springer, Heidelberg, Germany.

Körner, Ch., Neumayer, M., Pelaez Menendez-Riedl, S., and Smeets-Scheel, A. (1989). Functional morphology of mountain plants. *Flora* 182, 353–383.

Meyer, E., and Thaler, K. (1995). Animal diversity at high altitudes in the Austrian Central Alps. In *Arctic and Alpine Biodiversity: Patterns, Causes and Ecosystem Consequences* (F. S. Chapin III and Ch. Körner, eds.), pp. 97–108. Springer, Heidelberg, Germany.

Miehe, G. (1997). Alpine vegetation types of the central Himalaya. In *Polar and Alpine Tundra. Ecosystems of the World* 3 (F. E. Wielgolaski, ed.), pp. 161–184. Elsevier, Amsterdam, The Netherlands.

Rundel, P. W., Smith, A. P., and Meinzer, F. C. (1994). *Tropical Alpine Environments.* Cambridge University Press, Cambridge, United Kingdom.

Vuilleumier, F., and Monasterio, M. (1986). *High Altitude Tropical Biogeography.* Oxford University Press, New York.

AMAZON ECOSYSTEMS

Ghillean Prance
Royal Botanic Gardens, Kew

GLOSSARY

caatinga Within the Amazon region this is applied to open forest on white sand, which occurs in the basin of the Rio Negro. The term is also applied to the semidesert region of northeast Brazil; it is an indigenous word from the Tupi language meaning an open place.

capoeira Brazilian term for secondary forest on cleared ground.

cerrado The Brazilian term for the large area of savanna and savanna forests that are the dominant vegetation of the Planalto of Central Brazil.

igapó Vegetation periodically flooded by acidic black water or clear water rivers.

inselberg Granitic domes that rise above the forest in the older geological formations to the north and south of the alluvial plains of Amazonia and elsewhere.

llanos The term used in Colombia and Venezuela for the extensive grass-dominated, open savannas of the orinoco river region.

oligotrophic Term applied to water and soils that are particularly poor in nutrients, such as the weathered white sand soils of Amazonia.

rivers in Amazonia The three main types of rivers in the Amazon region are known as white, black, and clear water. White water rivers are muddy with much suspended sediment and are neutral or only slightly acidic. Black water rivers are dark because of dissolved tannic matter and are acidic (pH about 4). Clear water has neither mud nor humic matter and is usually slightly acidic.

tepui Venezuelan term for the sandstone table mountains of the Guayana Highland or Lost World region.

terra firme Brazilian term for areas that are above the level of periodic inundation by the rivers.

várzea Brazilian term for vegetation periodically flooded by white water rivers.

MOST PEOPLE THINK OF THE AMAZON REGION as being covered by a uniform green carpet of rain forest. This is far from reality, because Amazonia is a complex mosaic of ecosystems varying from typical tall rain forest to open grassland savannas to scrubby vegetation on white sand that resembles the chaparral of western California. Even when one type of vegetation is studied, many local variations occur depending on local edaphic conditions, the flooding regime, or rainfall. This great variety of ecosystems has been defined in different ways by different authors depending on their approach to the subject. For example, some authors have concentrated on the physiognomy and structure of the vegetation and others, such as Holdridge *et al.* (1971), on the life zones that define vegetation based on climatic factors such as rainfall, temperature, and altitude. In this account, the primary separation of ecosystems is based on floristics, that is, the individual species that make up an ecosystem, and on their physiognomy and structure.

In addition to the previously mentioned factors that account for the variety of ecosystems, the history of the region must also be taken into consideration. Over time the Amazon ecosystems have by no means been stable. The distribution of the different types of vegetation has fluctuated with changes in worldwide climate. The distribution and species composition of forest and nonforest biomes changed continuously during the Pleistocene and earlier epochs. What today is continuous forest was once broken up into isolated blocks interspersed with deciduous forest and savanna ecosystems that later coalesced when the climate again became favorable for forest. These changes have added to the overall complexity of the current Amazon ecosystems. The principal vegetation formations are described here.

I. NONFLOODED TERRA FIRME ECOSYSTEMS

The primary division of Amazonian ecosystems is between those that are subject to periodic flooding by the seasonal rise and fall of river water levels and those that are above the flood level. In Brazil the nonflooded land is known as terra firme, and since this term is widely applied in the literature about Amazonian vegetation it is also used here. Within the Amazon region there are forests and savannas on both terra firme and floodplain.

A. Rain Forest on Terra Firme

This is the single most widespread vegetation type of Amazonia and covers approximately 50 percent of the region. Rain forest occurs where there is heavy rainfall, usually over 2000 mm/yr, that is not markedly seasonal. Some of the other types of vegetation described here occur in areas with a marked dry season and/or with less rainfall, for example, on the southern fringes of Amazonia where transition forest and semideciduous forest occur.

The rain forest on terra firme is characterized by its dense closed canopy at about 25–35 m above the forest floor. Above the canopy a number of tall emergent tree species rise up to about 50 m. The dense canopy means that little light (about 3 percent) penetrates to the forest floor and the herbs and shrubs growing there are adapted to living in low light. The periodic sunflecks are of considerable importance to those species and to the seedlings of trees. Tree species range from light demanders that grow quickly where light is abundant, but that survive poorly in shade, to shade-tolerant species whose seedlings can survive and grow in low light.

Rain forest on terra firme is most notable for its amazing diversity of tree species. There have now been numerous inventories of trees in different parts of Amazonia. The highest number of trees of 10 cm diameter or more was recorded in western Amazonia at Cuyabeno in Ecuador, where Valencia *et al.* (1994) found 307 tree

TABLE I

Results of Some One-Hectare Inventory Plots in Amazonian Rain Forest Where Trees of 10 cm Diameter or More Were Counted

Site	No. of species	No. of trees	Habitat
Cuyabeno, Ecuador	307	693	Aseasonal rain forest
Yanomono, Peru	283 (+17 lianas)	583	Aseasonal rain forest
Cocha Cashu, Peru	189	650	Seasonal rain forest
Xingu River, Brazil	162	567	Seasonal rain forest
San Carlos de Rio Negro	83	744	Aseasonal forest on poor oxisol
Beni, Bolivia	94	649	Seasonal forest

species on a single hectare. Inventories have listed from 81 to 307 species depending on the soil, rainfall, and the seasonality of the climate (Table I). When the total number of woody species in a hectare is counted, there can be as many as 550 species.

The floristic composition is not uniform and some species have very local distributions. As a result, the lowland forest of Amazonia has been divided into a number of distinct phytogeographic regions based on the distribution of species. The first important subdivision of the region was offered by Ducke and Black (1953): this was later modified slightly by Prance (1977). The seven major phytogeographic regions can be explained by the history of the region and by the present-day climate. For example, the Atlantic coastal region has a wet almost aseasonal climate, but farther inland around Santarém, Brazil, there is a region with a much drier climate and a strong dry season. Farther to the west near the Andes in Peru, the climate is much wetter and less seasonal. Each of these regions falls into different phytogeographic zones.

Figure 1 shows the phytogeographic regions with the distributions of some plant species superimposed. While some species have extremely restricted local ranges and help to define the different regions, others are widespread and occur throughout the entire region, for example, *Caryocar glabrum, Licania heteromorpha,* and *Parinari excelsa*. The distribution of individual species is also dependent on altitude, soil, drainage, and topography, as well as climatic factors. There is not much structural or physiognomic difference in terra firme forest between the different phytogeographic regions, but the canopy height can vary considerably from one place to another.

Between 1972 and 1983, the RADAMBRASIL project surveyed the entire Brazilian Amazon using side-scanning radar. It distinguished two main types of forest on terra firme based on topography: forest in the lowlands on relatively flat ground and hill forest on undulating land between 250 and 700 m. The dense lowland forest occurs mainly on the area of Tertiary and Quaternary sediments between the major rivers on land below 250 m. This forest usually has a canopy at about 30 m and frequent emergent trees up to 50 m, for example, the Brazil nut (*Bertholletia excelsa*) or the angelim pedra (*Dinizia excelsa*), the largest tree in the forest. The un-

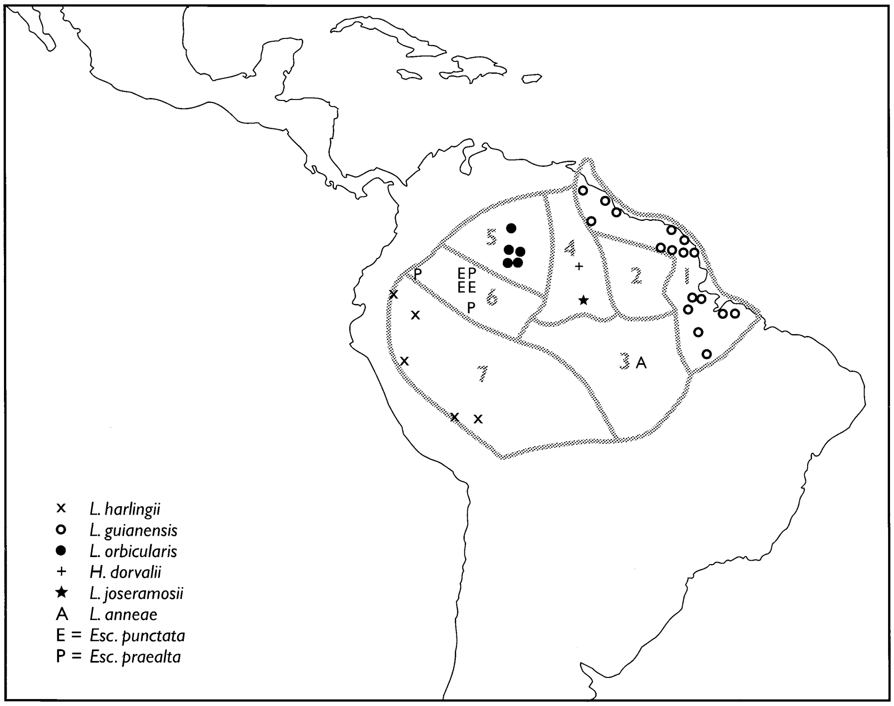


FIGURE 1 Map showing the phytogeographic divisions of the Amazon rain forest as defined by Prance (1977). The symbols represent the distribution of some species of trees typical to each region. *L.* = *Licania* (Chrysobalanaceae), *H.* = *Hirtella* (Chrysobalanaceae), *Esc.* = *Eschweilera* (Lecythidaceae).

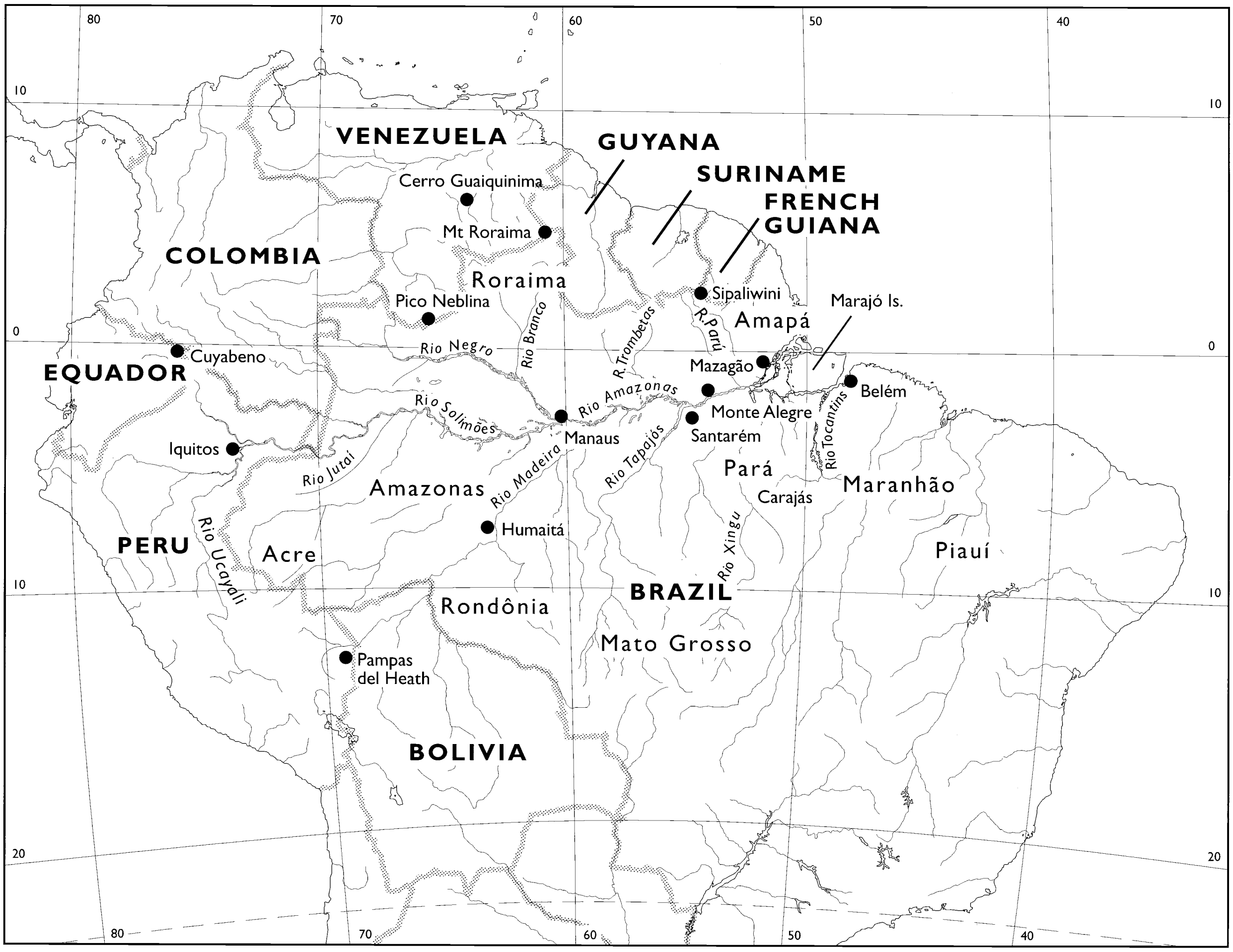


FIGURE 2 Map of northern South America, showing principal cities and rivers and localities mentioned in text.

derstory has relatively few shrubs and consists mainly of young individuals of canopy tree species. Above 250 m the lowland forest tends to be replaced by hill forest, which has a lower canopy and its emergents reach to about 35 m tall. The canopy is more open and consequently the understory is often much thicker and has a greater diversity of shrubs and herbs. The physiognomy of hill forest is far more varied than that in lowland forest, and there are often marked differences between slopes and valleys in both physiognomy and species composition, adding to the total species diversity of the region.

Most trees of the terra firme forest have large heavy fruit or seeds that drop to the ground. After this initial dispersal by gravity they are often carried farther by secondary dispersal agents such as rodents and other terrestrial animals. In contrast to savanna and riverine vegetation, most terra firme forest species are not adapted to long-distance dispersal. This must affect the rate of expansion of the forest into potentially colonizable habitats following climate change or human-caused forest destruction. The various species of primates in the canopy are important dispersers of seeds as well as seed predators.

B. White Sand Oligotrophic Formations

The edaphic factor that has the most striking effect on vegetation on terra firme is white sand soil. This habitat is scattered throughout the region, but especially between the Rio Branco and the Rio Negro in Brazil. These areas are characterized as podzols, or leached pure white sand, known as regosols in Brazilian soil terminology. The origin of these sands is either from the erosion of sandstone mountains in the northern part of Amazonia or from uplifted former river or sea beaches throughout the region. This soil is nutrient deficient and, as a consequence, has a distinctive lower vegetation that varies from an open savannalike formation to low forest. In the Old World tropics this is called heath forest, but a variety of local terms have been applied in the Neotropics, such as Amazonian caatinga, campina, campinarana, chavascal, and charravascal. The commonest term for campina forest in Amazonia is caatinga, but this leads to confusion because the same term is used for the semidesert thorn-scrub of northeastern Brazil.

The largest area of campina forest in the upper Rio Negro region has a canopy at about 18 m. The trees tend to be of a much smaller diameter than in terra firme forest and the branching is tortuous and does not form a closed canopy. As a result, the ground vegetation is abundant and rich and it contains many endemic species, especially in the families Bromeliaceae, Marantaceae, and Rapateaceae. There are also abundant epiphytic Orchidaceae, Araceae, and pteridophytes on the tortuous branches of the trees.

In addition to the almost continuous area of campina forest in the upper Rio Negro, there are many small isolated patches of campina especially in the lower Rio Negro around Manaus. These form small islands of campina surrounded by forest on terra firme. Some of the species that dominate these campina islands are *Aldina heterophylla, Humiria balsamifera,* and *Pradosia schomburgkiana.* These campina islands have trees with particularly tortuous branches that are heavily loaded with epiphytes. Besides the epiphyte families mentioned earlier, Gesneriaceae and *Peperomia* in the Piperaceae are common and they support ant gardens where ants nest among the roots and form a rich layer of detritus. In some campinas there are open areas where bare sand occurs that is often covered by the blue-green alga *Stigonema tomentosum,* as well as small islands of shrubs and low trees with abundant lichens around them. There is good evidence from the presence of pottery shards that these open areas were cleared by indigenous peoples about 800 years ago. The recolonization by vegetation has been extremely slow because of the lack of nutrients in the soil, and so many open patches still remain today.

A study of a campina island near Manaus showed that three-fourths of the species of trees and shrubs are adapted to long-range dispersal by birds, bats, or wind. Therefore, campina species tend to have a much greater capacity for long-distance dispersal than species of the terra firme forest.

White sand areas occur in many other places throughout Amazonia. One large area is that of Serra do Cachimbo on the border between the Brazilian states of Pará and Mato Grosso. This area has a similar physiognomy to the campinas of the Rio Negro, but the species composition differs considerably and there are a number of species that are endemic to the sands of Cachimbo. To the west of this region there are white sand areas in Amazonian Peru and Colombia and in the Brazilian state of Acre. In the Guianas, white sand areas are abundant and are covered by forest or by open savanna. Dense forests on white sand in Guyana are dominated by the legume species *Eperua falcata* and two other species of *Eperua* and are termed wallaba forest. To the south of Amazonia there are extensive areas of white sand vegetation in the Chapada de Parecís in Rondônia, where many of the same genera occur as in the Rio Negro region, for example, *Abolboda, Clusia, Humiria, Paepalanthus, Retyniphyllum, Ternstroemia,* and *Xyris*.

The other white sand formation that occurs in Amazonia is called restinga. This low scrub habitat along the seacoast of Brazil is found on sand dunes beyond the influence of salt water and inland from beaches or mangrove forest. Within Amazonia, restinga is not extensive but it extends southward around the coast to southern Brazil. Restinga is a rather uniform scrub characterized by the presence of *Chrysobalanus icaco,* a species that also occurs in similar habitats in West Africa. Common herbs include the ubiquitous *Ipomoea pes-caprae* and the sedge *Bulbostylis capillaris;* shrubs include *Hibiscus tiliaceus* and *Manilkara triflora.*

C. Dry Semideciduous Southern Fringe Forests and Cerradão

This formation occurs predominantly in southeastern Amazonia on the border with the cerrado (the savanna) of the Central Brazilian Plateau. This type of dry forest also extends into the cerrado region, where it is termed cerradão. In this region the climate is much more seasonal, drier, and less humid. As a result, the trees lose some of their leaves in the dry season to make the forest semideciduous in marked contrast to the evergreen terra firme forest. Dry forests occur in small patches and do not occupy an extensive area. A small amount of dry forest also occurs in northern Amazonia in Roraima state and in southern Guyana. The dry forests really form part of the transition zone between Amazonian high forests and regions of savanna and are interspersed with the various types of transition ecosystems described in the following section.

The semideciduous forest is not rich in endemic species. Some of the common species that are widespread and characteristic of this formation include *Bow-*

dichia virgiloides, Centrostigma macrophyllum, Combretum leprosum, Erythrina ulei, Lafoensia pacari, Magonia glabrescens, Attalea speciosa, Physocalymma scaberrimum, and *Vochysia haenkiana.*

A study of an area of this type of forest in southern Amazonia at Sarare in Mato Grosso found that it contained many species typical of Amazonian terra firme, such as *Enterolobium schomburgkii, Hernandia guianensis, Parkia* spp., *Schizolobium amazonicum, Simaruba amara,* and *Spondias lutea,* mixed with others more characteristic of the forests of São Paulo, such as *Aspidosperma, Myroxylon,* and *Poeppigia procera.* There are also a number of species in this belt of transition forest that do not occur in central Amazonia within the dense forest region. The best known of these species is mahogany, *Swietenia macrophylla,* which occurs around the western and southern fringes of Amazonia. The genus *Myroxylon* is another example of this type of distribution. The other region that contributes to the semideciduous ecosystem is the cerrado region of central Brazil. Species such as *Apuleia leiocarpa, Erythrina ulei, Plathymenia foliosa,* and *Physocalymma scaberrimum* are typical cerrado species that extend their ranges into the semideciduous forest.

II. TRANSITION ECOSYSTEMS

The transition forests form a belt of vegetation between the terra firme forests of Amazonia and the savannas of Roraima and Rupununi to the north and the cerrado of central Brazil to the south.

Transition forests mostly occur in regions with a strongly seasonal climate, but the boundary between them and the tall rain forest is not a clear-cut line. Rain forests extend farther into areas with seasonal climates that are on soils where the periodic water stress is compensated for by good water retention. This can occur because of either the physical properties of the soil or the proximity of rivers. Conversely, seasonal forests can extend into areas with little dry season in places that easily become dry, such as ridge crests and coarsely textured soils, and also on oligotrophic soils. Thus the superimposition of a climate map on that of vegetation will not necessarily correspond well and the forest boundaries are reticulate. The rain forest to seasonal forest ecotone still remains one of the least-studied ecosystems of Amazonia, yet it is in this boundary where most deforestation is occurring because of its accessibility to the developed south of Brazil and because the soil there tends to be more fertile than under evergreen forests. It is also easier to farm in a seasonal region when there is a time to burn the cut forest and to dry crops. In addition to the semideciduous forest, there are various, more specialized formations that tend to be dominated by one or a few species only. The three principal types of transition forest are described in the following.

A. Babassu Palm Forest

Babassu is a species of palm, *Attalea speciosa,* that dominates the transition forests of the southeastern fringe of Amazonia in the Brazilian states of Maranhão, Tocantins, and Pará. In a babassu palm forest, the trees are widely spaced and do not form a closed canopy. The tree height is from 10 to 25 m and many pure stands of *Attalea* occur. As a result, species diversity is extremely low and there are few endemics. *Attalea speciosa* is a fire-resistant species and this is probably the reason for its abundance in a transition region to savanna. The fires that occur naturally in the cerrado probably help to reduce the competition from other species and allow babassu to dominate. The extent of babassu forest has probably been increased by deforestation both historically by indigenous peoples and more recently by clearance for farming. Babassu occurs as a component of semideciduous forest, and when the latter is cleared, it is often one of the few species that survives the burning, hence the recent increase in area covered by babassu forest. Babassu is also an extremely useful species and oil is extracted commercially from the kernel of the fruit. The woody endocarp is used to make charcoal or as a fuel for factories, and the outer mesocarp is used as an edible flour.

B. Liana Forest

As the name implies, in a liana forest woody vines or lianas predominate. It is open forest with well-spaced trees that are often completely entwined by lianas. In Brazil this formation is known as cipoal or mata de cipó (cipó being the Brazilian word for liana). The tree species in liana forest are some of the commoner species of terra firme forest, but they occur at a much lower diversity. Babassu palm also occurs frequently in liana forest, but there is a far greater quantity of lianas here, particularly in the plant families Bignoniaceae, Dilleniaceae, Leguminosae, Malpighiaceae, and Menispermaceae. To walk through this forest, one has to cut one's way through a thicket of tangled liana stems.

Liana forest is especially abundant in the area between the Rios Tapajós, Xingu, and Tocantins in southern Pará, Brazil, but it also occurs in small patches in Roraima state in the north. The exact reason for the occurrence of this type of forest has not yet been established. One study showed that the soil under liana forest did not differ from that of nearby terra firme forest. It is thought to occur in regions where the forest has been disturbed by either human intervention or natural climate cycles.

Some of the typical Amazonian tree species that occur in liana forest include *Astronium gracile, A. lecointei,* the Brazil nut (*Bertholletia excelsa*), *Elizabetha paraensis,* and *Sapium marmieri.*

C. Bamboo Forests of Western Amazonia

Bamboo forest is an open type of transition forest that occurs mainly in southwestern Amazonia in Amazonian Peru, Bolivia, and the Brazilian state of Acre. In Acre, bamboo forest is abundant and covers a large area. In this forest various species of bamboo (*Bambusa* subgenus *Guadua* and the genus *Merostachys*) are extremely frequent in both the understory and the canopy, where they reach up to 30 m. Bamboo forest contains relatively few tree species and those that do occur are mostly common species of the terra firme forest. The bamboos are interspersed with trees, which lend support to their arching branches. By branching the bamboos spread far over the forest canopy and can easily be picked out in aerial photographs. Botanist Thomas Soderstom measured a bamboo culm of 29.77 m in length with a side branch of 9.69 m and a secondary branch that was a further 4.96 m. The belt of transition forest that extends across the southern fringe of Amazonia is dominated by bamboos to the west, lianas in the central part, and babassu palms to the east.

III. FLOODPLAIN ECOSYSTEMS

Almost 10 percent of the Amazon region is subject to periodic flooding due to the rise and fall of river levels in most of the region and backup of the rivers from high tides in the delta region. River levels are very seasonal and the magnitude of change is considerable. In central Amazonia around Manaus, the Rio Negro and Rio Solimões undergo an average annual change in water level of 10 m, and the difference between extremely high flood years and the extreme low level is 15 m. The forests beside the rivers are flooded to a depth of several meters for five to seven months each year. The two main flooded ecosystems are those of the white water rivers called várzea and those of the black and clear water rivers called igapó.

A. Várzea Forest

Várzea forest is formed on soil that is flooded by the alluvial-rich white water rivers. The soils are therefore much more fertile than those of the terra firme because of the annual deposit of alluvial matter. For this reason, much of the várzea forest has been replaced by agricultural systems that are more sustainable than those on terra firme. Seasonal várzea is forest that is periodically flooded by the seasonal rise and fall of river level, and tidal várzea is flooded twice a day by fresh water as a result of backup caused by the tide in the Atlantic Ocean. There are few differences between seasonal and tidal várzea and most plant species can occur in both types.

Várzea forest is tall and is physiognomically quite similar to terra firme forest, but it is much less diverse in species. Buttressed trunks are common in várzea forest trees, and lianas are also common. The várzea often has a rich understory with a large number of species of Zingiberaceae, Marantaceae, and Heliconiaceae. In upper Amazonia the várzea forests are much richer in species composition than those of lower Amazonia. Várzea forests tend to have high riverbanks that form natural levées. Some of the typical species are *Buchenavia macrophylla, Ceiba pentandra* (the kapok tree), *Crataeva benthamii, Gustavia hexapetala, Hevea brasiliensis* (the rubber tree), *Hura crepitans, Piranhea trifoliata, Rheedia brasiliensis,* and *Virola surinamensis,* as well as numerous species of palms such as *Astrocaryum jauari.*

In the lower Amazon, where the river is very wide, especially between Itacoatiara in Amazonas and Monte Alegre in Pará, the várzea forest is often accompanied by large meadows of grassland known as canarana (*Echinochloa spectabile* being the dominant species). This region has narrow bands of forest on higher ground beside the rivers and extensive grass meadows on the lower, more flooded ground farther away from the river margins; the lowest areas form lakes. When the waters recede and the lakes dry up, the open ground quickly becomes covered with grass, increasing the area of the meadows. The arboreal species that are most common in this type of várzea are *Bombax munguba, Calycophyllum spruceanum, Crataeva benthamii, Cordia tetrandra, Pithecelobium multiflorum, Muntingia calabura,* and *Salix martiana* var. *humboltiana.* The latter species of willow occurs mainly on muddy beaches that are in the process of formation.

B. Igapó

The black water and clear water rivers form a very different ecosystem than do the white water rivers. In this case there is little or no sediment and the water is acid and nutrient poor. In black water rivers the water contains much humic matter. The rivers flow from either sandy or rocky terrain. The largest black water river is the Rio Negro and the three major clear water rivers are the Rios Tapajós, Tocantins, and Xingu.

In these rivers in the dry season the banks are usually gradually sloping sandy beaches with trees on them, but in the rainy season they are transformed into an inundated igapó forest where only the canopy is above water. Characteristic species include *Alchornea castaniifolia, Couepia paraensis, Exellodendron barbatum, Eschweilera tenuifolia, Neoxythece elegans,* and *Virola elongata.* Two common palms of the igapó of the Rio Negro are the piassaba, *Leopoldinia piassaba* and the closely related *Leopoldinia pulchra.*

The physiognomy of the igapó forest is usually rather different from that of the terra firme and várzea forests. This forest is of a lower stature and the trees are more scattered. Many species are confined to the igapó forest but local endemism is rare. One interesting local endemic, *Polygonanthus amazonicus,* occurs on the sandy beaches near Maués in Pará, and it is still of uncertain plant classification.

Near Manaus, where the large black water Rio Negro and the white water Amazon (called the Rio Solimões above Manaus) merge, there are extensive areas of inundation forest that are influenced by both black water and white water. A recent study of this area showed that the floodplain forest contains a mixture of species that are characteristic of várzea and igapó as well as many of the species that are common to both ecosystems, such as *Astrocaryum jauari, Calophyllum brasiliense, Eschweilera albiflora, Gustavia augusta, Hevea spruceana, Parinari excelsa,* and *Vatairea guianensis.*

The seeds of both várzea and igapó species are often light and have many different flotation mechanisms, such as spongy tissue, oil bodies, hollow areas, or a light mesocarp. In some cases it is the seed that floats, and in others the entire fruit floats to ensure dispersal by the water currents. In the rubber tree the seeds float, and in the case of *Montrichardia arborescens,* an aquatic aroid, the whole fruit floats.

Other riverside trees have a fleshy fruit that is often colored green. These are the ichthyochorous or fish-dispersed species. When the fruit of these trees falls into the river in the flood season, fish can be seen grabbing them rapidly. The local people know that they can catch these fruit-eating fish by using tree fruit as bait on their hooks. For example, the fruit of *Alchornea castaniifolia, Carapa guianensis,* and *Astrocaryum jauari* are frequently used as bait. The seeds of most species of the igapó and várzea forest are dispersed either by water or by fish. Some of the shrubs that grow on riverbanks, such as *Myrciaria dubia* or the camu camu, *Alchornea castaniifolia,* and *Couepia uiti,* produce fruit as the river level rises. The fruit become submerged and are then readily available to their fish dispersal agents, who pluck them from the trees.

C. Lakes, Swamps, and Other Aquatic Systems

1. Lakes

Not all of the flooded ground is covered by the tall, species diverse, várzea and igapó forests. There are many lakes scattered throughout Amazonia and some of them are extensive, such as Lago Manacupuru in Amazonas and Lago Grande in Pará. In the upper Amazon in Peru and in the upper reaches of some of the main Amazon tributaries, oxbow lakes are also common.

The lakes of Amazonia have an extensive macrophyte flora. White water lakes beside the main Amazon River that are cut off from river flow in the dry season are the habitat of the world's largest water lily, *Victoria amazonica.* This spectacular plant is pollinated by scarab beetles that the lily traps inside its flowers for 24 hours. There are many species of Pontederiaceae, including the water hyacinth *Eichhornia crassipes,* which has become a noxious weed in many other parts of the tropics. Another common floating aquatic in the same family is *Reussia rotundifolia.*

In most lakes the aquatic flora can be divided into the free-floating species, those that are rooted to the bottom of the lake, and those growing around the margin of the lake. In the free-floating group there are several aquatic ferns such as *Azolla, Salvinia, Ceratopteris pterioides,* and *Marsilia polycarpa.* Others include a floating spurge, *Phyllanthus fluitans,* a legume, *Neptunia oleracea,* and *Ludwigia helminthorriza* in the willowherb family or Onagraceae. Some of the rooted aquatics are *Limnocharis flava, Pacourina edulis, Caperonia castaneifolia, Aeschynomene rudis, Ludwigia decurrens,* and *Sphenoclea zeylanica.* Growing around the lake margin are common species such as *Cassia occidentalis, Cyperus mutisii, Paspalum fasciculatum, Hyptis parkeri,* and *Diodea kuntzei.* When the lake level rises, large mats of floating vegetation, mainly the grasses *Panicum* and *Echinochloa,* are washed out into the river and can be seen floating downstream.

2. Permanent Swamps and Buritizal

In a few places in Amazonia the water never drains and permanent swamp occurs. The permanent swamp forest has a low species diversity and covers only a small area of the region. The soil under the forest is a eutrophic humic gley and in some areas of dystrophic humic gley palm swamp occurs. This is usually dominated by species of *Mauritia* or *Euterpe.* There are several extensive areas of *M. flexuosa* palm swamp that are locally called buritizal. These occur in eastern Amazonia in the state of Maranhão, in central Amazonia in the basin of the Rio Jutaí, and in the west near Iquitos, Peru. Since almost all parts of both *Mauritia* and *Euterpe* are useful, these oligarchic or one-species forests have great potential for sustainable use. *Euterpe oleracea* occurs in large quantities in the swamp forests of Marajó Island in the Amazon delta, where it is harvested extensively for both its fruit and as a source of palm hearts or palmito. *Mauritia flexuosa* also occurs in the gallery forests of the Roraima savanna and in várzea forest and is not confined to the permanent swamp forest.

3. Pirizal

Pirizal refers to a type of vegetation that occurs in small restricted areas of eastern Amazonia. A pirizal is a shallow lake or pond with stagnant water and a large number of rooted plants that emerge above the water. The commonest species are the large sedge, *Cyperus giganteus,* and *Thalia geniculata,* an aquatic member of the Marantaceae. There are also a number of typical floating aquatics such as species of *Cabomba, Eichhornia, Limnathemum, Nymphaea, Sagittaria,* and *Salvinia.* The *Mauritia* palm often grows around the margins. The pirizals occur as enclaves in the midst of dense forest, mainly in the state of Amapá in an area along the coast, especially in the region of Mazagão. There are also sedge-dominated communities in Amapá called cariazal. These are dominated by *Diplasia karataefolia,* the local name of which is cariá.

D. Mangrove Forest

The mangrove forest of Amazonia occurs as a narrow littoral belt in the coastal area that is subject to salt water inundation by tidal movements. Amazonian mangrove forests are species poor and have a rather uniform formation. The commonest species is the red mangrove, *Rhizophora mangle,* which occurs in the saltiest water. *Rhizophora. racemosa* occurs in less salty brackish water, and the intermediate hybrid species *R. harrisonii* occurs in the area in between. All three species are found in the mangroves of the Amazon coast.

The other common species of this formation are *Avicennia nitida,* the white mangrove, *Laguncularia racemosa,* and *Conocarpus erectus. Avicennia* grows inland for a considerable distance upriver into the zone of fresh water, whereas all three species of *Rhizophora* do not extend beyond the influence of salt water. In areas of clay salt beaches the small grass *Spartina brasiliensis* plays an important role in stabilizing the soil as it occurs in thick dense carpets. Some of the species that are common around the fringes of Amazonian mangrove are *Annona palustris, Hibiscus tiliaceus, Pithecellobium cochleatum,* and *Pterocarpus officinalis.* The Atlantic coastal mangrove forests of South America are species poor in comparison to those along the Pacific coast.

IV. AMAZONIAN SAVANNAS

A. Nonflooded Savannas and Cerrado

Scattered throughout the rain forest region of Amazonia are a number of patches of open grassland or savanna. Some of these cover small areas and others are quite extensive. The largest ones are the Roraima–Rupununi savanna in Guyana and northern Brazil, which covers 54,000 km^2, the Sipaliwini savanna on the border between Brazil and Suriname, the Pampas del Heath in Peru, and the Humaitá savannas in the south of Amazonas state. Amazonia is bordered to the south by the largest savanna area of South America, the cerrado of the Central Brazilian Plateau. There are also banks of cerrado that extend into the rain forest region of southern Amazonia, but the isolated Amazonia savannas are a rather different formation from the cerrado. The Amazonian savannas have less local endemism and species diversity than the cerrado, few suffrutices (woody plants with underground branches and short aerial shoots which die back after fires and resprout), and less tortuously branched trees, indicating that fire is not as strong a factor as it is in the drier cerrado.

The cerrado region of central Brazil has an extremely well defined dry season, the air humidity is often low, and the soils are very deep. As a consequence, roots are also deep and are specialized to reach water at considerable depth when the upper layers of soil are dry. When the topsoil is scraped off by a tractor, immediately there is a large amount of sprouting from vegetative organisms such as xylopodia. In contrast, in Amazonian savannas the roots are more superficial and there are not many vegetative parts in the soil to allow vegetative reproduction. Also, the climate in Amazonia is more humid and the relative humidity of the air is higher. Furthermore, in Amazonian savannas there is never such a dense arboreal cover as in the more closed types of savanna, whereas in the cerrados, dense closed vegetation (cerradão) is common and is similar to the dry semideciduous forest.

The two ecosystems have many species in common, for example, *Antonia ovata, Byrsonima verbascifolia, Curatella americana, Hancornia speciosa, Palicourea rigida, Qualea grandiflora, Salvertia convalarriodora,* and *Tabebuia caraiba.* The wide distribution of these and other species is evidence of a historically more continuous distribution of savannas during previous periods of drier and cooler climate. However, it has been shown that many savanna species have a much greater capacity for long-distance dispersal than rain forest trees. Bird- and wind-dispersed diaspores are common in savanna ecosystems, enabling distribution into isolated islands of savanna.

Most of the larger areas of savanna, such as those of Roraima, the savannas of the Rio Paru region of Pará state, and the extensive savannas around Santarém, also in Pará, occur within the drier and more seasonal regions of Amazonia. A belt of drier seasonal climate extends from Roraima to Santarém with under 2000 mm of rainfall and a strong dry season. Other savannas, especially some of the smaller patches, occur for edaphic reasons, such as an underlying hardpan.

There are a large number of savannas in Amazonian Venezuela and these were subdivided into three types by Otto Huber (1982): (1) grassy llanos and llanos-type savannas confined to the north of the region, (2) grassy inundated savannas of the Manapiare–Parucito basin, and (3) the true Amazonian savannas of the central and western part of Amazonas Territory. Unlike many of the Brazilian savannas, the Amazonian savannas of Venezuela have a high rate of endemism. They are mainly of edaphic, pre-Quaternary origin, whereas the llanos and inundated savannas are relict areas from the Pleistocene arid periods. The savannas of Amazonia are therefore not just one single ecosystem.

Other savannas also vary greatly in their botanical diversity and degree of endemism. The Roraima, Sipaliwini, and Humaitá savannas have few endemic species, whereas a few savannas have a rich and diverse flora. Most notable are those of the Rio Cururu region of Pará, which occur over sandstone. The sandstone savannas of the Guayana region of Venezuela are also richer in endemism than those of central Amazonia. Savannas that occur over special edaphic conditions are generally much richer than savannas occurring in terrain with similar conditions to those of the surrounding forest. These edaphic savannas are probably older and more stable over time.

Along the northern fringes of Amazonia, other extensive savannas are the llanos of Colombia and the contiguous Gran Sabana of Venezuela. There are also extensive upland savannas in the Departments of San Martín and Madre de Dios in Peru. Overall, savanna covers approximately 15 percent of the Amazon region and so it adds considerable species diversity. Many of the savanna ecosystems are under threat by overuse for farming and agriculture. For example, much of the cerrado region is being destroyed to plant soy and the Roraima–Rupununi savannas are being intensively burned by cattle ranchers. In the dry El Niño year of 1998, the fires spread into the rain forest, causing extensive destruction.

B. Flooded Savannas of Eastern Amazonia

In the delta region of Amazonia in the state of Amapá and in the eastern half of Marajó Island, there are large areas of inundated savanna or campos de várzea as they are called locally. Other inundated savannas occur farther inland in Pará, especially between the Rios Xingu and Tapajós, and in the basin of the Rio Madeira in Amazonas, but these are much smaller than those of Marajó and Amapá. There are many species of grass in the flooded savannas, such as *Eragrostis hypnoides, E. glomerata, Parathera prostrata, Paspalum orbiculatum,* and *P. guianense.* But the main difference between flooded and nonflooded savannas is the greater frequency of sedges (Cyperaceae) such as *Cyperus giganteus* in the wettest areas and *Cyperus luzulae, C. ferax,* and *Scleria geniculata* in the less flooded areas. Shrubs are less abundant in flooded savanna, but in areas that are grazed by cattle *Artemesia artemisiifolia* becomes so abundant that it can completely dominate an area because it is unpalatable to cattle. *Ipomoea fistulosa,* a shrub that is toxic to cattle, also increases on grazed land for the same reason. Other species characteristic of flooded savannas include *Aeschynomene sensitiva, Alternanthera philoxeroides, Capironia fistula, Cassia reticulata* (known locally as mata pasto or pasture killer because of its tendency to invade pasture), *Clitoria triquetrum, Justicia obtusifolia, Mimosa pigra, Montrichardia linifolia, Phaseolus lineatus, Rhabdadenia macrostoma, Sesbania exasperata,* and *Thalia geniculata.*

In the flood season, various floating aquatic species, such as *Ceratopteris pteridioides, Eichhornia azurea, Neptunia oleracea, Pistia stratiotes,* and *Salvinia radula,* multiply with great rapidity.

V. MONTANE FRINGING ECOSYSTEMS

Around the northern and southern edges of Amazonia, on the older pre-Cambrian formations of the Guiana and Brazilian crystallic shields, there are a number of hills and mountains, and to the west Amazonia is bordered by the Andes Mountains. These upland areas, which enclose the lowland Amazon basin on three sides, add considerably to the biological diversity of the Amazon ecosystem. Small changes in altitude or topography can cause significant changes in the vegetation type, physiognomy, species composition, and climate.

A. The Sandstone Tepuis of the Guayana Highland

An area with a complex mixture of vegetation types is the Guayana Highland, which is dominated by dramatic sandstone table mountains or tepuis. The mountain slopes are covered by lower and upper montane forest and often there is an abrupt change to open savanna and swamp formations on the summits of the mountains because of the sheer sandstone cliff faces. The highest mountains have less forest on their summits and more open vegetation except for gallery forest along the streams. The lower mountains have considerable areas of forest on their summits. For example, on Cerro Guaiquinima in Venezuela, one of the largest sandstone tepuis, 40 percent of the summit is covered by forest. Both tall forest and relatively low dense forest, as well as intermediate types, occur on Guaiquinima. There is a strong relationship in species content between the forest on the summit of this and other tepuis and the forest of the neighboring lowlands. Species endemism is high on the sandstone tops of the Guayana Highlands, and many genera have a series of closely related species with one occurring on each of the larger mountains, for example, the pitcher plant genus *Heliamphora* or the shrub *Tepuianthus*.

Two tepuis, Tepequém and Araca, occur within the Brazilian Amazon region, and Mounts Roraima and Neblina both lie on the Brazilian border with Venezuela. The latter is the highest tepui, reaching 3045 m in altitude. The Guayana Highland region is a complex mosaic of tepuis interspersed with rain forest or savanna depending on the locality.

B. Granite Inselbergs

Scattered throughout the Guianas and northern Amazonian Brazil are a series of granitic outcrops or domes that rise above the rain forest to a height of 300–800 m. In some cases these hills are covered by dense forest (e.g., Palunlouiméempeu and Mitraka), but most of them are characteristic inselbergs with low scrub forest and open areas of exposed rocks. These are obviously islands of a special type of vegetation surrounded by rain forest. The tops of the inselbergs are well drained and become very dry in the dry season, and so the vegetation is often dominated by sclerophyllous plants or species with other adaptations to drought, such as the orchid *Cyrtopodium andersonii* with large pseudobulbs that store water or various cacti (*Epiphyllum* and *Melocactus*). The dominant shrub is usually a species of *Clusia* and other terrestrial orchids include *Epidendrum nocturnum* and *Encyclia ionosma,* and the bromeliad *Pitcairnia geyskesii* is also common. Many of the plant species on the inselbergs of the Guianas are confined to that habitat. There has been considerable speciation in adaptation to the summit of inselbergs with their arid dry season conditions and very humid rainy season conditions. However, in contrast to tepuis, there has been much less speciation between the different inselbergs, and local endemics are rare.

C. Low Hills within Amazonia

In addition to the tepuis and inselbergs scattered throughout the Amazon region are several lower hills that can have distinct vegetation. These are more frequent to the north of the Amazon River and near the Brazilian Shield region and are mostly outliers of the ancient shields.

Small changes in altitude can have extremely important effects on vegetation type, physiognomy, species composition, and climate. For example, outside Amazonia, small low hills in the arid northeast of Brazil tend to accumulate cloud and therefore are covered by tall moist forest called brejo. The brejo forests contain many Amazonian species. The different effects between larger and smaller mountains (known as the Massenerhebung effect), as well as latitude, on the vegetation type are quite striking in the Neotropics. Consequently, there are many small patches of montane or cloud forest on small outcrops throughout the lowland region. Also the altitudinal limits of the different forest types vary considerably depending on the effects of local climate, soil, and latitude. Some of the species and genera that occur on the higher mountains also occur on these areas, for example, *Retinophyllum* and *Pagamea* in the Rubiaceae.

D. The Andes Foothills

The Amazon region is bordered on the west by the Andes, which rise abruptly out of the lowland rain forest. Unlike the tepuis and inselbergs, the Andes are a relatively recently formed mountain chain and have a distinct zonation of the vegetation according to altitude. The forests of the eastern slopes fall into the ecosystems of Amazonia and add considerable variety of both vegetation types and species.

The lowland forest extends up the mountains to between 700 and 1000 m. Above that altitude it is replaced first by lower montane, and then by montane vegetation types. The montane forests can generally be divided into three zones: lower montane forest, upper montane forest, and subalpine forest.

Lower montane forest begins between 700 and 1200 m. It is quite similar to lowland rain forest, but a large number of species drop out and are replaced by more upland species. This forest tends to be lower in stature than that of the surrounding lowland area, and it has fewer woody vines and fewer buttressed trees. Vascular epiphytes are common and mesophyll leaf types are predominant. Many of the lowland taxa persist into this zone, such as species of *Licania* (Chrysobalanaceae) and *Eschweilera* (Lecythidaceae), but a number of distinctly highland elements also enter the lower montane forest, for example, in the Colombian Andes, *Brunellia comocladifolia, Alchornea bogotensis,* and *Cinchona calycina.* Palms such as *Euterpe purpurea* and *Wettinia cladospadyx* are quite abundant, and woody vines include *Anomospermum occidentale* and species of *Passiflora* and *Paullinia.* This zone typically extends upward to between 1800 to 2400 m depending on latitude and local conditions.

The upper montane forest begins at 1800–2400 m and may extend in places up to 3400 m. It is usually of lower stature than the lower montane forest, with a predominance of microphyllous trees. Vascular epiphytes are still common and woody lianas are rarer. An increasing number of species characteristic of higher altitude enter the flora, for example, *Brunellia occidentalis, Weinmannia balbisiana, Symplocos pichindensis,* and the vines *Hydrangea peruviana* and *Liabum megacephalum.*

The subalpine forest, including the dwarf elfin forests, is of frequent occurrence in the Andes and extends locally into Central America. It is microphyllous or, at its altitudinal extreme, nanophyllous or dwarf-leaved. Few vascular epiphytes and climbers occur, but there are abundant bryophytes and lichens. This formation may extend up to 3800 m in some places, above which it is replaced by the herbaceous alpine formations such as páramo and puna, which are not discussed here. The subalpine forest has few predominantly lowland genera but many that are characteristic of the highlands, for example, species of *Weinmannia* (Cunoniaceae) and *Gynoxys* (Asteraceae), *Brunellia* (Brunelliaceae), *Clusia* (Clusiaceae), *Befaria* (Ericaceae), *Miconia* (Melastomataceae), and *Rhamnus* (Rhamnaceae).

A most interesting type of dry forest occurs in small isolated dry valleys that are scattered along the Andes due to local weather conditions They are significant as a link between the larger arid areas, and they were probably more nearly connected during the dry periods of the Pleistocene and early Holocene. Important genera in these valleys include *Acacia, Bursera, Caesalpinia, Cercidium, Prosopis* (*P. chilensis* and *P. limensis*), and many Cactaceae.

E. Campo Rupestre

The campos rupestres, or open formations on rocks, are generally confused with open savanna and the orchard savanna on terra firme, but they are quite different physiognomically and floristically. They develop on rocks and in rocky terrain, because they suffer drought in the dry season despite the equatorial humid climate, and because there is no possibility for water retention and all rainfall runs off immediately. Typical species of the cerrado do not occur here, such as *Curatella americana, Hancornia speciosa, Salvertia convallariodora,* and *Qualea grandiflora.* Instead there are various species of *Byrsonima, Clusia, Norantea,* and *Vellozia,* certain Bromeliaceae (*Bromelia, Dyckea,* and others), and many Eriocaulaceae (*Paepalanthus* and *Syngonanthus*). Lichens are also frequent and cover many of the rocks.

Campos rupestres are also quite common in central Brazil, where they are often confused with cerrados. For example, Serra do Cipó in Minas Gerais, much cited in the literature because of its interesting landscape and large number of Velloziaceae, is mostly composed of campo rupestre. In Amazonia, Serra do Cachimbo, in the southwest of Pará, has an area of campo rupestre that covers about 16,000 km^2. Another large patch is the campo of the Rio Cururú, in the Tapajós River basin. At Ariramba, on the Rio Trombetas, both savanna and campo rupestre occur.

The canga or ironstone vegetation, which occurs on Serra dos Carajás over the iron deposit, is also a special form of campo rupestre. On this formation, various extra-Amazonian species are common, such as *Pilocarpus microphyllus,* which occurs also in the northeastern states of Piauí and Maranhâo; the extra-Amazonian genus *Callisthene* is represented by *C. microphyllus* of central Brazil, which occurs only within Amazonia on Serra Carajás. Also found on Carajás is the curious and beautiful treelet *Norantea goyasensis*. The branches of this thick-barked treelet sometimes lengthen and turn it into a vine.

The Amazonian montane forests also show a strong affinity to campo rupestre and some could even be classified as floresta rupestre (i.e., forest on rock). In the mountain areas there has been a great accumulation of organic matter because, although the forests sometimes endure short dry periods, they are maintained by the high air humidity and mist.

VI. SECONDARY FOREST

Unfortunately, secondary forest is of increasing importance in Amazonia owing to the amount of felling of primary forest that is taking place. Obviously the species that now dominate secondary areas had a natural distribution prior to the occurrence of man-made secondary forest. A few secondary forest species are found throughout the tropics, and others grow in open spaces beside rivers, on old landslides, in gaps caused by tree falls or dead trees, or where forest has been felled by severe storms. Because the seeds and stumps have not been destroyed by fire, regenerated vegetation of naturally cleared areas is different from that in areas burned by man. Secondary forest species do not generally occur in the transition regions between forest and savanna. Modern vegetation maps show an increasingly large area of a new man-made vegetation and also of agricultural areas, but these are not discussed here. Some of the characteristic genera of secondary forests are *Cecropia, Byrsonima,* and *Vismia*. In Brazil, secondary forest is known by the term capoeira.

VII. CONCLUSIONS

This summary of the principal vegetational formations that occur in Amazonia shows the considerable diversity that accounts for the approximately 30,000 species of vascular plants within the region. Many of the ecosystems described here could be subdivided further into many different local communities. A knowledge of this ecosystem diversity and the distribution of species is an essential tool for the conservation and sustainable use of the region. To preserve the diversity of Amazonia, it is vital to conserve areas of each different type of vegetation because of the number of unique species of plants and animals that each ecosystem contains.

See Also the Following Articles

RAINFOREST ECOSYSTEMS • SOUTH AMERICAN ECOSYTEMS • SOUTHERN SOUTH AMERICA AND OTHER AUSTRAL ACOSYTEMS

Bibliography

Daly, D., and G. T. Prance. (1989). Brazilian Amazon. In *Floristic Inventory of Tropical Countries* (D. G. Campbell and H. D. Hammond, eds.), pp. 401–426. The New York Botanical Garden, New York.

Ducke, A., and G. A. Black. (1953). Phytogeographical notes on the Brazilian Amazon. *Anais Acad. Bras. Cienc.* **25**(1), 146.

Gentry, A. H. (1982). Patterns of neotropical species diversity. *Evolutionary Biol.* **15**, 1–84.

Gentry, A. H. (1988). Tree species richness of upper Amazonian forests. *Proc. U.S. Natl. Acad. Sci.* **85**, 156–159.

Gentry, A. H. (1989). Northwest South America (Colombia, Ecuador and Peru). In *Floristic Inventory of Tropical Countries* (D. G. Campbell and H. D. Hammond, eds.), pp. 391–400. The New York Botanical Garden, New York.

Gentry, A. H. (ed.). (1990). *Four Neotropical Forests*. Yale University Press, New Haven, Connecticut.

Holdridge, L. R. (1972). *Forest Environments in Tropical Life Zones; A Pilot Study*. Pergamon Press, Oxford/New York.

Huber, O. (1982). Significance of savanna vegetation in the Amazon Territory of Venezuela. In *Biological Diversification in the Tropics* (G. T. Prance, ed.), pp. 221–244. Columbia University Press, New York.

Hueck, K. (1966). *Die Walder Südamerikas*. Gustav Fischer Verlag, Stuttgart, Germany.

Pires, J. M. (1984). The Amazonia forest. In *The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and Its Basin* (H. Sioli, ed.), pp. 581–602. Dr. W. Junk, Dordrecht, Netherlands.

Pires, J. M., and G. T. Prance. (1985). The vegetation types of the Brazilian Amazon. In *Key Environments: Amazonia* (G. T. Prance and T. E. Lovejoy, eds.), pp. 109–145. Pergamon Press, Oxford, United Kingdom.

Prance, G. T. (1977). The phytogeographic subdivisions of Amazonia and their influence on the selection of biological reserves. In *Extinction Is Forever* (G. T. Prance and T. S. Elias, eds.), pp. 195–212. The New York Botanical Garden, New York.

Prance, G. T. (1979). Notes on the vegetation of Amazonia III. The terminology of Amazon forest types subject to inundation. *Brittonia* **31**, 26–38.

Prance, G. T., and T. E. Lovejoy (eds.). (1985). *Key Environments: Amazonia*. Pergamon Press, Oxford, United Kingdom.

Prance, G. T. (1987). Vegetation. In *Biogeography and Quaternary History in Tropical America* (T. C. Whitmore and G. T. Prance, eds.), pp. 28–44. Oxford Monographs on Biogeography 3. Clarendon Press, Oxford, United Kingdom.

RADAMBRASIL. Levantamento de recursos naturais Vegetação. Vol. 8, pp. 307–403. Min. das Minas e Energia, Dept. Nacional da Produção Mineral Projeto Radambrasil, Rio de Janeiro.

Sioli, H. (ed.). (1984). *The Amazon: Limnology and Landscape Ecology of a Mighty Tropical River and Its Basin*. Dr. W. Junk, Dordrecht, Netherlands.

Valencia, R., H. Balslev, and G. Paz Y Miño. (1994). High tree alpha-diversity in Amazonian Ecuador. *Biodiversity and Conservation* **3**, 21–28.

AMPHIBIANS, BIODIVERSITY OF

Ross A. Alford,* Stephen J. Richards,† and Keith R. McDonald†
*James Cook University; †Queensland Parks and Wildlife Service

GLOSSARY

anamniotic Eggs are not surrounded by the complex membranes that distinguish the amniotic eggs of reptiles, birds, and mammals.

cloaca The common chamber in which the reproductive, excretory, and digestive tracts of amphibians unite before exiting the body.

metapopulation A group of local populations among which individuals migrate relatively frequently; however, the rate of migration is slow enough that the populations fluctuate independently.

paedomorphosis Reproduction while retaining at least some larval characteristics.

tetrapods The terrestrial vertebrate classes amphibians, reptiles, birds, and mammals, so named because they primitively possess four legs.

AMPHIBIANS ARE TETRAPOD VERTEBRATES. They differ from the other tetrapods (the reptiles, birds, and mammals) in that their eggs are anamniotic; they are relatively simple and are enclosed in a jelly capsule.

I. THE EVOLUTIONARY HISTORY OF AMPHIBIANS

Amphibians first appeared during the Late Devonian, about 360 million years ago (Ma). There is a general consensus that all amphibians shared a common ancestor, a sarcopterygian (fleshy-finned) bony fish in the class Osteichthyes. All sarcopterygians have paired fins with limb-like bones, and many exhibit other anatomical features, such as lungs, that ally them with the tetrapods. It is not likely that either of the extant groups of sarcopterygians, the dipnoans (lungfishes) and crossopterygians (lobe-finned fishes), contained the ancestor of the amphibians. The earliest major radiation of terrestrial vertebrates occurred during the Carboniferous Period (ca. 335 Ma). By the end of the Triassic, about 200 Ma, nearly all of the large ancestral amphibians were extinct. The subclass Lissamphibia, the modern amphibians, appeared during the Triassic and is the only group that has survived to the present.

The ancestry of modern amphibians is poorly understood because there is a sparse fossil record linking primitive amphibians to the three modern orders. The earliest fossil Lissamphibian is *Triadobatrachus massinoti* from the early Triassic, about 230 Ma. *Triadobatrachus* is similar to the modern frogs but is not considered to belong to the order Anura. The Anura (frogs) and Gymnophiona (caecilians) appeared during the Early Jurassic (about 190 Ma), whereas the Caudata (salamanders) appeared in the Middle Jurassic, 170–150 Ma. Although the fossil evidence is sparse, phylogenetic analyses of shared derived morphological characters and of molecular characters strongly suggest that the Lissamphibia share a common early amphibian ancestor and that the Lissamphibia and the amniote tetrapods (reptiles, birds, and mammals) originated from different early amphibians. It is possible, however, that the Lissamphibia is a polyphyletic group, with one or more of the modern orders having an independent origin from the subclasses Temnospondylii or Microsauria.

II. HISTORICAL BIOGEOGRAPHY AND CURRENT DIVERSITY OF MODERN AMPHIBIANS

Modern amphibians are found in all continents except Antarctica. However, the three modern orders show disparate patterns of fossil and recent diversity that are associated with their different histories on the landmass of Pangaea and subsequently on landmasses associated with Gondwana and Laurasia. The evolution of salamanders is linked closely with landmasses derived from Laurasia. All fossil and living species occur in the Northern Hemisphere, with the exception of the lungless salamanders (family Plethodontidae) which invaded South America relatively recently. In contrast, the caecilians and most frogs are associated predominantly with the southern Gondwanan landmasses. The caecilians occur throughout the tropics, except in Madagascar and to the east of Wallace's Line in Australasia. They are the least diverse group of living amphibians, with five families and 165 species. The greatest diversity of species occurs in northern South America and central America. Frogs are the most widespread of the three orders. Although the breakup of Pangaea probably isolated salamanders on the Laurasian continents and many groups of frogs on the Gondwanan continents, several clades of frogs have dispersed to all continents and frogs are now absent only from Antarctica.

There are now more described species of amphibians than there are of mammals (Glaw and Kohler, 1998; Table I). More than 400 salamander species are currently known in 10 families, with the highest diversity found in North America. One family, the Plethodontidae of North and South America, contains more than half the known salamander species.

TABLE I

Richness of Modern Amphibian Taxa*

Taxon	Families	Genera	Species
Anura	25	334	4204
Caudata	10	61	411
Gymnophiona	6	33	165
Total	41	428	4780

* Based on Duellman 1993, Glaw and Kohler 1998.

The frogs are placed in 25 families, and more than 4200 species are known, 80% of which are in tropical regions. It is likely that at least several hundred species remain to be described. The greatest diversity of frogs occurs in South America, but other tropical areas such as Southeast Asia, New Guinea, Madagascar, and central Africa also have highly diverse frog faunas.

Four of the five most diverse frog families, the Hylidae, Ranidae, Bufonidae, and Microhylidae, are each found across several continents—distributions that reflect successful dispersal. The most diverse family is the Leptodactylidae, which is restricted to South and Central America and contains 22% of living frogs. It contains the genus *Eleutherodactylus*, which is the most diverse vertebrate genus with more than 550 species.

III. BASIC MORPHOLOGY AND FUNCTIONAL ANATOMY OF MODERN AMPHIBIANS

All modern amphibians have complex glandular skins and most lack scales. Their skins are kept moist by the secretions of mucous glands, whereas granular glands produce a variety of toxins that serve to deter predation and may also help to protect from microbial infections. They shed their skins periodically and usually consume them as they are shed. Their color patterns are produced by xanthophores, iridophores, and melanophores. Many species are able to alter the shape of pigment cells and the distribution of pigments within them and can rapidly change color.

The eyes of most species contain photoreceptors of several types, probably giving at least limited color vision, they can be covered by moveable eyelids, and they focus by moving the lens using ciliary muscles. Many frogs and salamanders have a wide binocular field of vision and have good depth perception. Many species use olfaction in the detection and capture of prey and have well-developed olfactory systems and Jacobson's organs. Larval amphibians and adults of aquatic species have lateral line systems. The amphibian auditory apparatus differs from those of other terrestrial vertebrates. Their middle ears have a columella bone (also found in reptiles and birds and modified in mammals) which functions in the reception of high-frequency sounds. A second bony element in the middle ear, the operculum, receives low-frequency vibrations from the air and substrate via a muscular connection to the pectoral girdle.

Most terrestrial amphibians respire (in part) with lungs and have largely separate pulmonary and systemic blood circulations. Their hearts typically have three chambers—two atria that receive blood from the body and the lungs and one ventricle that serves to pump blood to the lungs and the body. Although the ventricle is usually single, a combination of partial septa and the high viscosity of blood reduce mixing of the blood streams flowing to the lungs and the remainder of the body. Many species carry out a considerable fraction of gas exchange across the moist skin; for the lungless salamanders (family Plethodontidae) this is the primary mode of respiration.

The kidneys of amphibians function in maintaining water and ionic balance. In species with aquatic larval stages, most water and ion balance needs are reversed at metamorphosis; the freshwater larvae must cope with a hypo-osmotic environment, to which they lose ions and from which they gain water, whereas the terrestrial stages must cope with rates of evaporative water loss that are typically high because of their moist skins. Aquatic individuals tend to excrete dilute solutions of ammonia, whereas terrestrial animals excrete more concentrated solutions of urea, although no amphibian excretes fluid urine that is more concentrated than the blood plasma. Water uptake in almost all amphibians does not involve drinking but is accomplished by active or passive transport of water across the skin.

Amphibians are ectotherms, relying on the external environment as a source of body heat. All aquatic forms and many terrestrial forms are also poikilotherms or thermoconformers, with body temperatures that do not differ from their environments. Some species may control their body temperatures to some extent by selecting microhabitats that provide appropriate temperatures, and a few species of frogs periodically increase their body temperatures by basking in sunlight. The ability of most terrestrial amphibians to thermoregulate by basking is limited by their need to conserve water while maintaining a moist skin. Some species of frogs and salamanders can be active at temperatures approaching 0°C, whereas others can tolerate body temperatures well above 40°C. Some temperate species can survive exposure to temperatures below 0°C by supercooling. A few species can survive partial freezing by a combination of ice nucleating proteins that encourage freezing of extracellular fluids and high intracellular glucose concentrations that prevent intracellular water from freezing and keep cells from dehydrating.

A. Special Features of Caecilians

Caecilians are aquatic and burrowing animals that superficially resemble large earthworms. Adults range from approximately 10 to 150 cm in length. They have elongate bodies with distinct annuli, which are grooves delineating their body segments. They are limbless, and their tails are reduced or absent. Their eyes are reduced and are covered by skin. They are unique among the Lissamphibia in possessing dermal scales, which occur in the annuli of some species. Their skulls are heavily ossified and completely roofed. Caecilians possess a unique chemosensory organ, the tentacle, which extends a short distance from the surface of the head, emerging from a skull opening between the eyes and the nostrils.

B. Special Features of Salamanders

Salamanders are typically four-limbed animals with relatively long tails that superficially resemble lizards but lack epidermal scales and claws. Salamanders range from about 30 mm to 2 m in total length. Their limbs are relatively small and are reduced or lost in some terrestrial and aquatic species. Their skulls typically show the loss of many bony elements. Salamanders lack external ears and, with the exception of weak distress calls in some species, do not vocalize. The most diverse group of salamanders, the Plethodontidae of the Americas, is lungless. Salamanders of many groups exhibit various degrees of paedomorphosis. Salamander larvae are carnivorous and usually have well-developed external gills.

C. Special Features of Anurans

Anurans are invariably four-limbed, and terrestrial juveniles and adults completely lack true tails. Adults range from about 1 to 30 cm in length. Their hindlimbs and feet are greatly elongated. The radius and ulna of the forelimb and the tibia and fibula of the hindlimb are fused. There are no more than nine trunk vertebrae, and most modern frogs lack free ribs. The caudal vertebrae are fused into a rod-shaped urostyle that is associated with the elongate pelvis. Most of these features are probably the result of adaptation for light weight and strength for jumping and show interesting convergences with similar adaptations in birds. The lightweight skulls of frogs are large relative to their body size and lack many bones. All but one species lack teeth on the dentary bone of the lower jaw. The tongue is attached at the front of the mouth in most species and is flipped forward rapidly to capture prey. The ears of frogs follow the general amphibian model, with the addition of an external tympanic membrane in many species.

The skins of frogs depart from the usual amphibian pattern in several ways. Some species have additional types of glands: lipid glands that secrete lipids which reduce rates of evaporative water loss or breeding glands that produce sticky secretions which adhere the male to the female during amplexus. Skin lipids and other as yet poorly understood modifications of the skin allow some treefrogs to achieve rates of evaporative water loss as low as those of some lizards. Frogs also have distinct differences between their dorsal and ventral skins. The ventral skin usually has fewer mucous glands, and many terrestrial species have an area called the pelvic patch in which the skin is unusually thin and highly vascularized. Frogs sit with the pelvic patch in contact with moist substrates to increase the rate of water uptake. Burrowing desert frogs of several lineages form cocoons by repeatedly shedding their skins and retaining the shed layers to reduce evaporative water loss while they are buried.

Frogs have a variety of adaptations to structures other than the skin for maintaining water balance. Most species store copious quantities of dilute urine in the bladder when water is available and withdraw water from this pool to replace evaporative losses. Desert species and species that inhabit brackish water can allow high concentrations of urea to accumulate in the body fluids. A few species can excrete nitrogenous wastes as uric acid, which minimizes the water lost when they are excreted.

IV. REPRODUCTIVE BIOLOGY AND LIFE HISTORIES OF MODERN AMPHIBIANS

The "typical" life history of the Lissamphibia includes a complex life cycle in which eggs are deposited in freshwater habitats, where larvae grow and develop, eventually metamorphosing into terrestrial juveniles and leaving the water. Although this is a common pattern found in all three modern orders, it is far from universal. One of the major features that sets amphibians apart from the remainder of terrestrial vertebrates is their extremely wide range of life histories and modes of reproduction, many of which occur in all three extant orders. The presence of a wide range of life histories and reproductive modes suggests that the "typical" amphibian life cycle does not reflect a failure to adapt to the terrestrial environment but rather serves as an adaptation that allows female amphibians to produce very large numbers of eggs with a small investment in each and to exploit freshwater habitats for larval growth and development.

A. Caecilians

Limited observations have been made on the reproductive biology of caecilians and no information on courtship behavior is available. It appears that fertilization is internal via protrusion of the male cloacal wall. Although vocalizations have been reported for some species, it is not known if they are linked to reproduction.

Caecilians are oviparous or viviparous. In viviparous species gestation may take up to a year with reproduction occurring only every 2 years, and nutrition can be supplied to the young within the oviduct. Oviparous species lay terrestrial eggs but the larvae can be aquatic or can complete development within the egg. Oviparous caecilians produce more offspring than viviparous species. In oviparous species, parental care of eggs is common.

B. Salamanders

Many salamanders conform to the "typical" amphibian complex life cycle. Most produce aquatic eggs and larvae, which metamorphose into terrestrial juveniles. As in the frogs, there is a great deal of variation beyond this basic mode. Larval salamanders are relatively similar to adults; this has allowed species belonging to several groups to evolve the ability to reproduce while retaining larval characters, either facultatively or obligately.

An example of the complexity and variability of salamander life histories is the life history of red-spotted newts (*Notophthalmus viridescens*, family Salamandridae). Their life history begins with aquatic eggs which hatch into aquatic larvae. The larvae typically develop within one summer and metamorphose into a morphologically distinct terrestrial juvenile stage, the eft. The eft stage lasts from 1 to 8 years, depending on temperature and food availability. Efts undergo a second transformation into adults, which then return annually to water to breed. Larvae that encounter very favorable conditions and grow rapidly can retain some larval morphological characteristics, remain permanently aquatic, and reproduce as paedomorphic adults. Larvae that encounter slightly less favorable conditions can bypass the eft stage and metamorphose directly into the adult morphology.

Courtship in salamanders can be quite elaborate, incorporating chemical, visual, and tactile cues. During the breeding season males of aquatic species can develop enlarged fins and become more brightly colored. There may be other morphological changes to the head glands, tooth structure, musculature, and skin during the breeding season. Many salamanders have specialized glands that secrete compounds used as olfactory signals during courtship.

Salamanders have internal or external fertilization. External fertilization involves deposition of the sperm on the egg mass. Salamanders with internal fertilization either transfer sperm directly from male to female or exhibit a unique form of sperm transfer in which a bundle of sperm (the spermatophore) is deposited by males. The spermatophore is picked up in the female cloaca and stored in a special structure in the cloaca, the spermatheca. The stored sperm can remain viable until ovulation, which may occur from a few days to 2.5 years later. In live-bearing salamanders the sperm enters the oviduct, in which the eggs are internally fertilized.

Both male and female salamanders of some species exhibit parental care (Fig. 1A), although it is practiced mostly by terrestrial breeders and may serve to protect the eggs from predation, fungus infection, and desiccation. Adults attend clutches for up to 9 months. In a few species communal clutches are attended by aggrega-

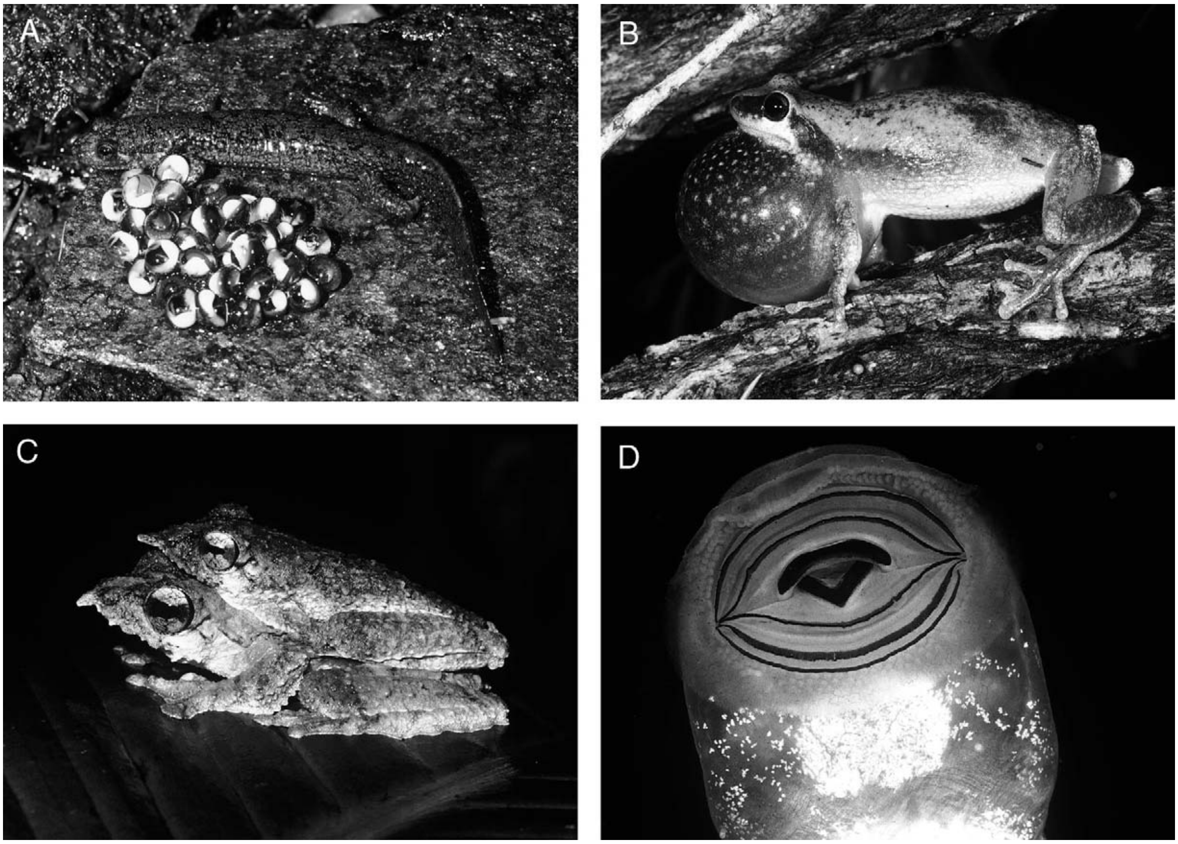


FIGURE 1 (A) A black-bellied salamander *Desmognathus quadromaculatus* guards its eggs [removed from (and replaced in) a small stream, Appalachian Mountains]. (B) A male Australian desert treefrog, *Litoria rubella*, produces an ear-splitting call with the aid of its balloon-like vocal sac. (C) Spike-nosed frogs, *Litoria prora*, from New Guinea, in amplexus. The function of the rostral (nose) spine is unknown. (D) A suctorial tadpole of the Australian treefrog *Litoria nyakalensis*. The upper and lower jaw sheaths, surrounded by keratinized "teeth" and oral papillae set in an oral disc, are clearly visible. These tadpoles use their large ventral oral discs as suckers to attach to rocks in fast-flowing rain forest streams (all photographs taken by and copyright by Stephen J. Richards). See also color insert, this volume.

tions of females, and eggs are fertilized by many different males.

C. Anurans

The most common reproductive strategy in frogs involves a complex life cycle with externally fertilized aquatic eggs which produce highly specialized larvae—tadpoles. The tadpoles grow and develop for some period in the water and then undergo a radical metamorphosis and emerge as terrestrial juvenile frogs. However, frogs have evolved a remarkable variety of reproductive strategies, most involving a trend toward removal of eggs and larvae from the aquatic environment. Pough *et al.* (1998) described 29 combinations of egg deposition sites, including still water and fast-flowing water, terrestrial nests, and on or in the body of the male or the female, and they also described tadpole development, ranging from the typical free-swimming feeding larva to direct development inside the oviduct. Because amphibian eggs lack a water-resistant shell, the greatest diversity of reproductive modes occurs in the humid tropics where eggs can survive for long periods in the permanently moist terrestrial environment. Complex behaviors including egg guarding and embryo transport (on the dorsum, in the mouth, and in *Rheobatrachus* species in the stomach), and unusual morphological structures such as skin pockets to provide protection for developing embryos, are associated with many reproductive modes. Several species bear live young, and at least one of these, *Nectophrynoides occidentalis*, actually provides oviductal nutrition to its developing embryos.

In tropical regions where conditions for reproduction are favorable throughout the year, breeding can be aseasonal, and females may lay multiple clutches in one year. In more temperate or high-altitude regions breeding is typically strongly seasonal, occurring only during short periods of the year when temperature and rainfall reach critical levels. Under these conditions females generally lay only one clutch of eggs each year.

Male frogs vocalize mainly to attract mates (Fig. 1B) and to advertise their presence and sometimes their status to other males. Calls are species-specific and each female's brain is tuned to respond only to males of her own species. In most frogs, males possess a single or double vocal sac which serves as a resonator and in at least some species as a sound radiator. Vocal sacs also conserve energy by allowing passive reinflation of the lungs as the vocal sac contracts after a call. Some male frogs, such as the Australian torrent frog (*Litoria nannotis*), lack a vocal sac but can still produce a surprisingly loud call. Environmental conditions such as temperature affect vocalizations. At colder temperatures notes and pulses are produced at slower rates, but the length of the call increases. The dominant frequencies in the calls of most frogs are lower than 5000 Hz, although those of some small species are higher. Within species, variation in the dominant frequencies, pulse rates, and durations of calls often reflects male body size; therefore, the call may indicate male quality as well as male location. Females of many species use these characteristics to choose their mates from among competing males. Males vocalize from species-specific locations which can be in water, on or beneath the ground, in vegetation from near ground level to high in trees, and even under water (several species including African clawed frogs, *Xenopus*).

Female frogs do not have a vocal sac and very few vocalize. Some female frogs produce a scream when distressed, and reproductively active females of some species call softly in response to male advertisement calls.

The posture of frogs during the fertilization of eggs is called amplexus; in most species this involves the male grasping the female from above (Fig. 1C). The exact posture adopted depends on the morphology and relative size of the male and female. The two most common positions involve the male grasping the female in front of the back legs (inguinal amplexus) or the front legs (axillary amplexus). Amplexus is aided in many species by specialized patches of skin called nuptial pads on the forelimbs of males. Pairs remain in amplexus while the male sheds sperm onto the eggs as they are released by the female. Fertilization in nearly all frogs is external but several species accomplish internal fertilization by cloacal apposition. The frog *Ascaphus truei*, commonly called the tailed frog, breeds in fast-flowing streams of the Northwest of North America and carries out internal fertilization using the "tail," which is actually an intromittent organ formed from an extension of the male's cloaca.

V. ECOLOGY AND FUNCTIONAL MORPHOLOGY OF LARVAL AMPHIBIANS

A. Caecilians

Relatively little is known about the larvae of caecilians. They are more similar to adults than are those of frogs or salamanders. Externally they closely resemble adults but have gill slits and fins. Free-living caecilian larvae have long external gills and a lateral line system. Their mouth and dentition resemble those of adults. They lack the tentacle organ that appears on the head of adults; this appears at metamorphosis.

B. Salamanders

Salamander larvae are much more similar to adults than are the tadpoles of frogs. Larval salamanders have external gills that are not completely covered by an operculum. Some embryonic salamanders have paired lateral projections from the head called balancers; in some species these persist for a short period following hatching. Most species possess well-developed fore- and hindlimbs through most of the larval period. Their bodies are laterally compressed compared to those of adults, and their tails are also relatively thinner and deeper. Their skins contain lateral line organs (neuromasts) and are thinner and less glandular than those of adults. Their dentition is different from adults, and their tongues are rudimentary. Their eyes lack lids. Larval salamanders are almost all carnivorous, usually feeding on zooplankton and larval insects. The larvae of larger species can also feed on small vertebrates. The larvae of some species have alternative morphologies; the typical morphology is usually a planktivorous carnivore, but when conditions are favorable some individuals develop relatively larger heads and more powerful jaws, adopting a "cannibal" morphology that allows them to prey on small vertebrates, including their siblings.

As in larval frogs, there is considerable variation among and within species in rates of larval growth and development. Both the minimum and the maximum rates for salamanders are slower than those for frogs. Salamanders can take from 6 weeks to 5 years to complete larval development. Within species, rates respond to both temperature and food availability, and salamanders have the additional option, apparently not available to frogs, of changing the relative rates of development of somatic and reproductive structures so that they mature sexually without losing all larval characters.

C. Anurans

The tadpole larvae of frogs are highly specialized for growth and development in the aquatic environment. They have an oval head–body region and a long tail, which is laterally compressed and includes a central area of musculature and dorsal and ventral fins of thin, lightly vascularized tissue. The tail is supported only by a notochord. Despite their very different body form, they swim and turn as rapidly and efficiently as fishes of similar body sizes. They feed using an elaborate pumping mechanism that is very different from the oral and branchial morphology of adults. This mechanism transports water though the mouth and pharyngeal cavity, where food particles are removed by branchial filters and entrapment in strands of mucus. Some tadpoles can remove particles as small as 0.126 μm from the water. Water is ejected through the nostrils of most species and through the spiracle, which is usually a single, tubular structure leading out of the opercular chamber and can be located midventrally or on the left side of the body. The unique mouthparts of tadpoles typically include an oral disc with transverse rows of keratinized "teeth" that are used to scrape particles into suspension. Keratinized sheaths on the jaws provide cutting and biting surfaces. The oral apparatus is variously modified and sometimes allows attachment to the substrate via suction (Fig. 1D). The relatively long, coiled intestine fills most of the body cavity. Tadpoles are typically thought of as microphagous herbivores that feed on algae and small parts of higher plants, but most species will feed on animal material when it is available. Tadpoles often scavenge on dead animals in the water and frequently prey on amphibian eggs.

Tadpoles often hatch with external gills, which are quickly covered over by a fold of epithelium, forming the opercular chamber. Before the opercular chamber forms, many species do not swim but attach to a substrate using adhesive organs located posterior to the mouth. Tadpoles lack limbs at hatching. The rear legs usually develop slowly, starting as limb buds at the posterior end of the body and developing over a long period. The forelimbs develop within the opercular chamber and are visible only after they erupt fully formed through the opercular wall at the onset of metamorphosis.

A typical pattern of tadpole growth and development would include 1 to a few days as a nonswimming hatchling with exposed gills followed by several weeks to months as a swimming and feeding larva. During this period, the tadpole grows dramatically; many species increase their body mass by hundreds of times and some by thousands. During this period, the hind-limbs begin to grow and slowly develop through the remainder of larval life.

Metamorphosis begins with the eruption of the forelimbs and involves drastic changes to all elements of the structure and function of the body. Reorganization of the mouth and digestive tract allows a switch from larval filter feeding to adult carnivory. The tail fins and musculature are broken down and reabsorbed. Most elements of the chondrocranium are reshaped and realigned, and the branchial apparatus ceases to be a support for gills and takes on a role as support for the adult tongue. The external and middle ears form, and in all but the aquatic frogs of the family Pipidae the lateral line system disappears. The eye changes in size and structure, the photopigments of the rods change from porphyropsin to rhodopsin, and eyelids appear. The axial skeleton is reorganized and many elements are ossified. The lungs, which develop and begin to function during the larval period of many species, enlarge and take on their role as a major respiratory structure. The complexity of the skin increases, with the epidermis increasing from two layers of cells to five or six, many with specialized functions. The kidney, which in tadpoles is relatively simple and excretes excess water and ammonia, becomes more complex to serve its new function of conserving water and excreting urea. The gonads differentiate at about the time of metamorphosis.

Rates of growth and development of tadpoles are typically highly variable within species, responding to environmental temperature, food availability, and the density of tadpoles of their own and other species. Many species that inhabit unpredictable environments, such as temporary ponds, can have larval periods from 2 weeks or less up to months. Some species regularly spend 1 year or more (maximum 3 years) as larvae. The interaction between rates of growth and development in tadpoles has produced a rich literature that examines how and why this interaction is controlled. In general, it appears that rates of growth control rates of development during the earlier part of the free-living tadpole stage (Wilbur and Collins, 1973). Larvae that are growing slowly develop more slowly than fast growers, and changes in growth rate caused by changes in environmental conditions are mirrored by alterations in developmental rate. Fast-growing tadpoles tend to metamorphose both earlier and at larger sizes than slower growing individuals. Very slow growers appear to regulate their developmental rate so that they metamorphose near a species-specific minimum size after a larval period that may vary greatly in length. This flexibility in relative rates of growth and development is lost late in the larval period at a point that may vary among taxa. Theory suggests that the regulation of these rates ultimately responds to natural selection acting in a way that depends on the relative rates of growth and survival in aquatic and terrestrial environments. These ideas have been applied to life-history transitions in many nonamphibian taxa, including fishes, insects, and plants.

Because frogs typically deposit large numbers of eggs during relatively short breeding seasons, densities of tadpoles are often high. The success or failure of a cohort of tadpoles in a typical temporary freshwater habitat depends on a highly complex and unpredictable set of factors, including the density of tadpoles of their own species and other species, which control the degree of intraspecific and interspecific competition for food and space; the number and species of predators present; and the duration of the aquatic phase of the habitat. Competition and predation are both controlled by the choice of time and place of breeding by adult frogs and to some extent by microhabitat selection within habitats by tadpoles. The outcomes of species interactions involving tadpoles can be altered by changes in the timing of breeding, and microhabitat selection by tadpoles can depend on the species and sizes of other tadpoles present.

Tadpoles are preyed on by a wide variety of vertebrate and invertebrate predators, for whom they constitute a valuable resource. Major predators include fishes, salamanders and salamander larvae, and the aquatic larvae of insects such as dragonflies, damselflies, and beetles. Vulnerability of tadpoles to predators typically decreases as the tadpoles grow and develop, and many tadpoles exhibit short-term behavioral responses to predators, such as decreasing their levels of activity or switching from midwater feeding to substrate feeding, that appear to decrease their vulnerability. Most of these responses also decrease the tadpoles' rates of growth and development, leading to trade-offs that have been explored by behavioral ecologists.

VI. BEHAVIOR AND ECOLOGY OF POSTLARVAL AMPHIBIANS

The ecological breadth of the three classes of modern amphibians is reflected in their geographical distributions (Savage, 1973). The caecilians are restricted by both thermal and water requirements to relatively low latitudes and elevations, and they do not occur where mean annual temperatures are less than about 12°C or total annual precipitation is less than about 1000–2000 mm. The salamanders have less restrictive ecological requirements and are distributed across a broader range of habitats, occurring from low to high latitudes and elevations. Their major limitation is clearly moisture; they do not occur in areas that have prolonged dry seasons and only rarely in areas with total rainfalls less than 1000 mm per annum. The anurans as a group can tolerate wide ranges of both temperature and water availability, and they occur at all but the highest latitudes and elevations and in all but the driest deserts.

Almost all adult amphibians are carnivores that ingest invertebrate and vertebrate prey small enough to be swallowed whole. Amphibians are mostly either sit-and-wait predators or active foragers, but none engage in cursorial pursuit of prey. Sit-and-wait predators locate their prey primarily using vision, whereas active foragers often use olfaction.

Salamanders use a variety of methods for prey capture. Many larvae, and adults of aquatic taxa, are suction feeders, using rapid depression of the floor of the throat to pull in water and prey. Terrestrial salamanders usually feed by extending their large, fleshy tongues, which adhere to prey and pull it into their mouths. The tongue of salamanders is attached at the base and is protruded and elongated by muscles and fluid tension. The degree of attachment of the tongue and the length to which it is protruded vary among taxa. Captured prey are ingested whole. Terrestrial anurans also capture prey by protruding their tongues. These are usually attached near the front of the lower jaw and are protruded by literally flipping them forward and downward. Prey that adhere to the tongue are drawn into the mouth and swallowed whole.

Amphibians generally appear to grow throughout their lives, but rates of growth decline drastically after reproductive maturity is attained. In captivity, many species can live for decades, and even in nature some species live for extended periods. In general, amphibian life spans appear to be limited more by environmental hazards than by aging and senescence.

Many species of frogs and salamanders occupy relatively small home ranges during the nonbreeding season and migrate to breeding habitats for reproduction. The nonbreeding home range is aggressively defended by some species. Some species return to their natal ponds to breed, whereas others may simply migrate to a suitable body of water. In temperate regions, some species may also regularly migrate to overwintering sites. Amphibians use celestial navigation, light polarization, and the earth's magnetic field as means of orientation during migrations. There is considerable movement among local populations in many species, particularly by anurans. Some of this movement is dispersal by juveniles, but some is due to longer range movements by terrestrial adults. At least one species of frog, *Bufo marinus*, can move great distances over relatively short periods. The range of the introduced Australian populations of this species has expanded by approximately 30 km per year for an extended period, and the animals that arrive first in new habitat are adults. Detailed studies of the local movements of adult *B. marinus* suggest that they are nomadic. Many mark-recapture studies of frogs have found high rates of disappearance of marked frogs and of the appearance of unmarked animals, suggesting that adults of many species may at least occasionally disperse to new areas.

The relatively high rates of migration between local populations found in many amphibian species suggest that groups of local populations often form metapopulations. It is important to recognize this because the dynamics of metapopulations are governed by different factors than those of local populations, and the conservation of metapopulations requires a different approach than the conservation of local populations.

Because amphibians typically produce relatively large numbers of relatively small eggs, their populations can increase rapidly in size when reproduction is successful. It is likely that populations of most amphibians normally fluctuate fairly widely over time (Alford and Richards, 1999). These normal fluctuations may include fluctuations of local populations to extinction, followed by relatively rapid recolonization by immigrants from adjacent local populations belonging to the same metapopulation. It is likely that for many species the persistence of most local populations has little bearing on whether the regional metapopulation will persist. However, there may be a few critical local populations that either serve as reliable sources of colonists or serve as stepping stones for migration between more widely distributed local populations. Identifying these local populations and conserving them will be crucial for ensuring the long-term persistence of many species.

VII. AMPHIBIAN CONSERVATION

A. Human Uses of Amphibians

Amphibians have featured prominently in many human cultures through stories, song, and poetry. In urban areas they are frequently found coexisting successfully with humans in parks and garden ponds. Frogs are an important source of protein in some subsistence cultures. In affluent countries, frogs are imported for consumption in gourmet restaurants. Hundreds of millions of frogs have been exported from Southeast Asia and the Indian sub-continent, resulting in increasing insect pest populations. Frogs have also become model organisms in ecological, embryological, physiological, and genetic research.

Amphibian skin contains a wide variety of chemicals, including complex amines, alkaloids, and polypeptides, some of which have pharmacological properties. Some skin toxins are effective against amphibian bacterial and fungal infections, and the skin of the South American frog *Epipedobates tricolor* contains a constituent, epibatidine, that blocks pain 200 times more effectively than morphine. Poison dart frogs of the family Dendrobatidae harbor many exceptionally toxic skin compounds and one species, *Phyllobates terribilis*, contains sufficient toxin in a single frog to kill several adult humans. This toxin is smeared on darts used by the Choco Indians of western Colombia for hunting monkeys and other large game. There are connections between frog diets and the presence of skin toxins; some dendrobatid frogs with high toxin levels in the wild gradually lose their toxicity when held in captivity. Although indigenous cultures have recognized the toxic and medicinal properties of frog skin for centuries, their potential for development as medicines using scientific methods has only recently emerged as a significant field of research.

B. Amphibians as Components of Ecosystems

Amphibians form a vital link in many food chains. They represent the highest vertebrate biomass in some ecosystems and occupy an intermediate position in the food chain. Aquatic larval amphibians are herbivorous to omnivorous (Anura) or carnivorous (Caudata and Apoda) and are significant prey items for a wide variety of vertebrate and invertebrate predators. Tadpoles in lakes and ponds often reach extremely high densities and can have a significant impact on nutrient cycling within these aquatic habitats. Because anuran larvae feed on algae and other aquatic material, they play an important role in the transfer of plant energy to predators of tadpoles. Adult amphibians feed on a wide variety of live food. Most are generalists that consume any living creature smaller than their gape size. However, some have specialized to feed exclusively on narrow ranges of food items, such as worms, ants, and even snails. Predators of amphibian eggs, tadpoles, and adults include other amphibians, spiders, insects, mammals, birds, and reptiles (especially snakes). Because of their important role in ecosystems, population declines or extinctions of amphibians may have significant impacts beyond the amphibian species affected.

C. Amphibian Diversity and Levels of Threat

There are approximately 5000 amphibian species (Table I), and new species are being discovered every year. Amphibians occupy all continents except Antarctica, and they are found in habitats ranging from arid deserts and saline mangrove swamps to tropical rain forests and mountain peaks more than 4000 m high. More than 80% of amphibians are found in the tropics, with an estimated 44% of the world's amphibians in the tropical Americas. Many tropical regions, including New Guinea, South-east Asia, and parts of northern South America, have been inadequately surveyed, and there is no doubt that many previously unknown species will be discovered in these areas.

The 1996 International Union for the Conservation of Nature (IUCN) *Red Data Book* lists 5 amphibian species as extinct and 124 as threatened, which represents 25% of the species for which assessment of conservation status has been undertaken. However, this information is influenced to some degree by the research and assessment focus in affluent countries and may change substantially when comparable research is undertaken in other geographic regions. There is little doubt that extensive habitat loss in tropical regions is causing the extinction of poorly known and undiscovered species.

D. The Problem of Amphibian Population Declines

Well-documented declines and disappearances of amphibian species and entire suites of species occurred over wide areas in the 1980s and 1990s. All amphibian populations fluctuate, and assessing the significance of downward trends in amphibian populations has been difficult. However, the widespread loss of species and populations in a relatively short time frame, including dramatic extinctions of previously abundant species in protected areas such as national parks, is evidence that declines are real.

Many causes have been postulated for amphibian declines, including habitat loss and modification, predation, environmental toxicity, disease, immunosuppression, ultraviolet radiation, changes in climate or weather patterns, and combinations of these. No single cause has been identified and declines need to be assessed case by case. In some areas populations of a suite of species have declined while other ecologically similar species have not been affected. Populations of some species have declined at high altitudes but remained unaffected at low altitudes. Elucidating the causes of these declines remains a difficult and complex problem.

Determining the causes of declines can be difficult. In areas of extensive urbanization, such as Europe and North America, declines or extinction of amphibian populations or species through habitat loss have clearly occurred. However, understanding rapid population declines in relatively pristine montane rain forests is a more challenging problem. A major hindrance to our understanding of population declines is the lack of information about amphibian autecology; how populations behave, and the extent to which they operate as metapopulations. Experimental ecology aimed at testing hypotheses about population declines will be vital for identifying causal factors. The role of diseases in amphibians is poorly understood, and many diseases have only recently been documented. Although monitoring of populations over time is essential to understand population behavior, it will not identify the causes of amphibian population declines. The amphibian decline problem is currently the focus of much research effort, as reviewed by Alford and Richards (1999).

VIII. CONCLUSION: THE AMPHIBIAN SUCCESS STORY AND ITS FUTURE

Contrary to popular opinion, modern amphibians are not a relictual remnant of the ancestors of other tetrapods but are a highly successful group in their own right. There are more species of amphibians than there are of mammals. Modern amphibians occur in nearly all of Earth's terrestrial habitats, from within the Arctic Circle to tropical deserts. Groups of modern amphibians that need to conserve water have evolved impermeable skins, cocoons, and the ability to excrete uric acid. Groups that need to breed outside water have evolved a startling array of reproductive adaptations; amphibians have the widest range of reproductive modes of any tetrapods. These include aquatic eggs and larvae, many species with terrestrial eggs, and truly viviparous species in which the mother provides nutrition in addition to the yolk during development. This diversity indicates that the typical reproductive pattern, with aquatic eggs and larvae, must not represent a constraint that has limited their success. It probably represents a successful adaptation that allows the exploitation of aquatic habitats by terrestrial species and allows a much higher fecundity than is available to species that must provision their eggs with enough yolk for complete development.

The ability to respond to environmental challenges has allowed the modern amphibians to persist and flourish during and through periods of dominance of terrestrial habitats by other tetrapod groups. They have outlived the early dominant amphibians, several waves of dominance by reptiles including the dinosaurs, and many radiations and extinctions of mammals. The current concern regarding declines and disappearances of many amphibians is justified because it may be an early manifestation of a general crisis in biodiversity. It seems likely, though, that as long as terrestrial habitats continue to accommodate vertebrate life, some amphibians will persist.

See Also the Following Articles

BIRDS, BIODIVERSITY OF • FISH, BIODIVERSITY OF • MAMMALS, BIODIVERSITY OF • REPTILES, BIODIVERSITY OF

Bibliography

Alford, R. A., and Richards, S. J. (1999). Global amphibian declines: A problem in applied ecology. *Annu. Rev. Ecol. Syst.* **30**, 133–165.

Cogger, H. G., Zweifel, R. G., and Kirshner, D. (Eds.) (1998). *Encyclopedia of Reptiles and Amphibians*, 2nd ed. Academic Press, San Diego.

Duellman, W. E. (1993). Amphibian species of the world: Additions and corrections. University of Kansas Museum of Natural History Special Publication No. 21, pp. 1–372. University of Kansas, Lawrence.

Duellman, W. E., and Trueb, L. (1986). *Biology of Amphibians*. Johns Hopkins Univ. Press, Baltimore.

Glaw, F., and Kohler, J. (1998). Amphibian species diversity exceeds that of mammals. *Herpetol. Rev.* **29**, 11–12.

Heyer, W. R., Donnelly, M. A., McDiarmid, R. W., Hayek, L. C., and Foster, M. S. (Eds.) (1994). *Measuring and Monitoring Biological Diversity: Standard Methods for Amphibians*. Smithsonian Institution Press, Washington, DC.

McDiarmid, R. W., and Altig, R. (Eds.) (1999). *Tadpoles: The Biology of Anuran Larvae*. Univ. of Chicago Press, Chicago.

Pough, F. H., Andrews, R. M., Cadle, J. E., Crump, M. L., Savitzky, A. H., and Wells, K. D. (1998). *Herpetology*. Prentice Hall, Englewood Cliffs, NJ.

Savage, J. M. (1973). The geographic distribution of frogs: Patterns and predictions. In *Evolutionary Biology of the Anurans: Contemporary Research on Major Problems* (J. L. Vial, Ed.), pp. 315–445. Univ. of Missouri Press, Columbia.

Wilbur, H. M., and Collins, J. P. (1973). Ecological aspects of amphibian metamorphosis. *Science* **182**, 1305–1314.

Zug, G. R. (1993). *Herpetology: An Introductory Biology of Amphibians and Reptiles*. Academic Press, San Diego.

ANTARCTIC ECOSYSTEMS

Peter Convey
British Antarctic Survey, Natural Environment Research Council

GLOSSARY

ablation Direct transfer of water molecules from ice crystals to the vapor phase, without transition through the liquid phase.

anhydrobiosis Ability of certain organisms to lose all detectable water from their bodies under certain conditions, yet remain viable when subsequently rehydrated.

cryoturbation Process by which soil particles and stones are moved and mixed by the frequent formation and subsequent melting of ice crystals in the soil column.

cryptogam Plants without apparent reproductive organs; plants that reproduce by spores.

fellfield Dry, windswept habitat of cold regions comprising mineral soil, gravels, stones, and rock, dominated by cryptogams (especially mosses and lichens) and, outside the Antarctic, by compact cushion-forming phanerogams and short grasses.

microclimate The climate experienced over a scale of millimeters to centimeters, within which the majority of Antarctic terrestrial biota (animals, plants, and microbes) exist.

microhabitat The small-scale habitat within which an organism or community exists.

nunatak Isolated ice-free summit of a mountain whose bulk is permanently below the surface level of an ice sheet.

phanerogam Plants with conspicuous reproductive organs (flowers); plants that reproduce by seeds.

propagule Any part of an organism capable of developing into a new individual (e.g., seed, spore, gemma, lichen soredium, or egg).

sub-nivean Pertaining to the environment beneath a temporary, seasonal, or permanent covering of snow.

THE TERRESTRIAL ENVIRONMENTS OF ANTARCTICA are among the most extreme on Earth, challenging the very existence of life itself. This article will outline briefly the geological and biological history of the continent, discuss the conditions currently experienced, and then describe its recognized terrestrial biogeographical zones and extant biota. Major determinants of terrestrial biodiversity and ecosystem function will be discussed and a consideration given to natural and human-induced processes of ecosystem development and change.

I. ANTARCTICA

A. Introduction to the Continent

Antarctica is distinct among the Earth's continents, not least because it is the only one without a long-term history of human occupation and influence. It is a large continent, with greater surface area than Australia. However, 99.5% of the continental area is covered permanently by snow or ice. Terrestrial ice-free ecosystems are limited to exposed nunataks, cliffs, and seasonally exposed snow- and ice-free areas, with habitats generally being more extensive and exposure of longer duration on a ring of sub-Antarctic islands surrounding the continent. Ice-free terrestrial habitats are patchily distributed over the continent, and are best represented in coastal regions, particularly along the Antarctic Peninsula, and in the major mountain ranges inland. Many ice-free areas are necessarily small in extent and isolated from others by distances of tens to hundreds of kilometers. However, in the McMurdo Dry Valleys region of southern Victoria Land, ice-free areas of several hundred square kilometers exist. Antarctica is isolated from other continental land-masses by the 1000-km-wide Drake Passage south of Tierra del Fuego and by 4000–5000 km of the Southern Ocean that lies between it and Australia and South Africa.

Antarctica plays an important controlling role in the world climate, through its reciprocal influences on atmospheric and oceanic circulation patterns. For much of the year, Antarctica maintains a negative energy balance, losing heat energy to the atmosphere. This acts to maintain the continental ice cap, which on average is 2 km, and in places up to 4 km, deep. This ice mantle gives Antarctica the highest average altitude of any continent. It is also the coldest continent, having recorded the lowest surface temperature on Earth (−89.6°C), and is the windiest. Despite the fact that the vast continental ice cap develops from snow precipitation, large areas of Antarctica are classified as cold deserts, with extremely low precipitation rates (much of which is directly lost by ablation, rather than entering terrestrial ecosystems).

B. Tectonic, Glacial, and Biological History

Antarctica formed an integral element of the Gondwanan supercontinent. As these landmasses separated, Antarctica's last continental links were with Australia and South America. The connection with South America, via what was to become the Antarctic Peninsula, was lost approximately 25–30 million years ago. At this time, although lying at a high latitude, Antarctica was not glacierized and abundant fossil evidence indicates that it possessed a cool temperate fauna and flora closely related to those then present in South America and Australia/New Zealand. The opening of an ocean passage between Antarctica and South America allowed the development of circumpolar ocean currents, eventually effectively isolating Antarctica from lower latitude sources of heat energy and leading to gradual cooling and formation of the continental ice cap. The extent of glaciation has varied widely both geographically and over time during the last 20 million years, with evidence of local warmer periods as recently as 1–3 million years ago being sufficient for the development of areas of cool temperate rain forest dominated by southern beech (*Nothofagus* spp.).

Pleistocene (Ice Age) glaciation mirrored that experienced in the Northern Hemisphere. At its maximum, the continental ice cap was considerably thicker than at present, and the continent was surrounded by extensive floating ice shelves. The island groups of the South Shetlands, South Orkneys, and South Georgia were similarly the centers of large ice caps that extended tens of kilometers into the surrounding ocean. At this time, terrestrial habitats would probably have been limited to isolated inland nunataks; most, and probably all, current lowland (coastal) habitats would have been obliterated by ice. Bathymetric models suggest that, even allowing for lower sea levels, floating edges of ice shelves would not have been interspersed with exposed terrestrial habitats. There is glaciological evidence supporting wide fluctuations in glacial extent throughout the Pleistocene and even in the last few centuries, paralleling information available in the Northern Hemisphere. Unlike most northern polar latitudes, Antarctica is still in the process of exiting from Ice Age conditions.

C. Recognized Biogeographic Zones

Antarctic terrestrial environments include those found on the continent itself and also on a ring of Southern Ocean islands south of approximately 50° latitude (Fig. 1). In this area terrestrial ecosystems with widely varying characteristics are to be found and a number of systems of classification have been proposed (see R. I. L. Smith, in Laws, 1984; Longton, 1988), some applying only to the Antarctic and others attempting to relate similar zones in both Antarctic and Arctic polar latitudes. Currently, three biogeographical zones are recognized: the sub-Antarctic, the maritime Antarctic, and the continental (or frigid) Antarctic. Although elements

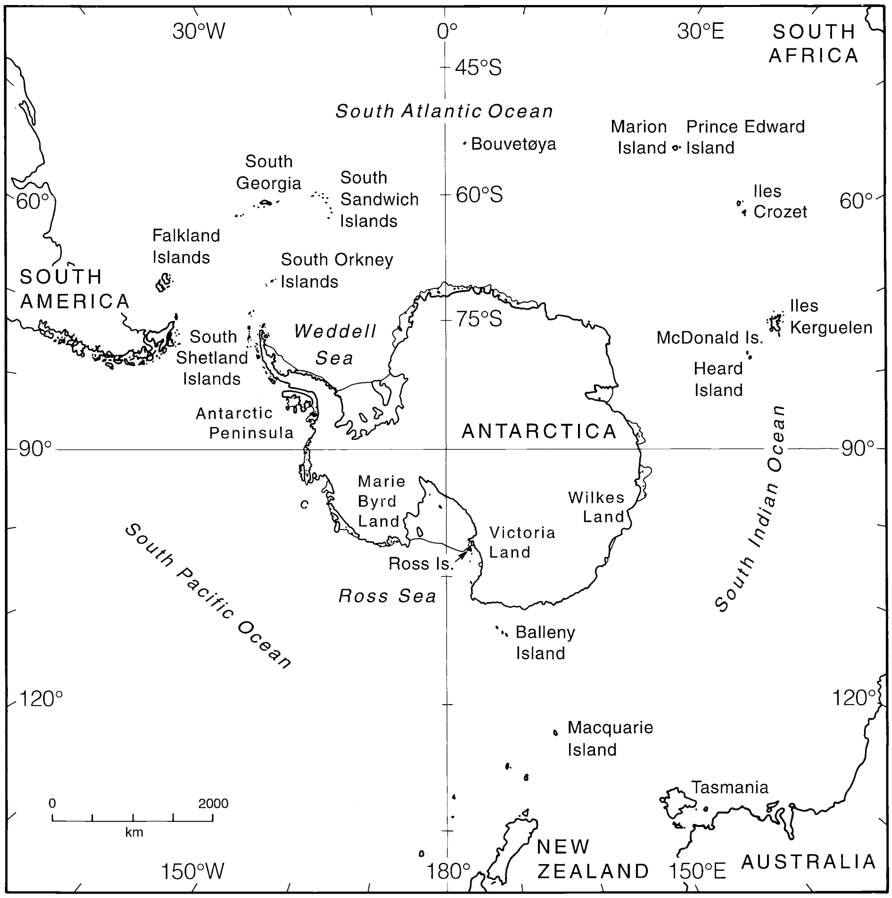


FIGURE 1 Map of Antarctica and associated Southern Ocean island groups, indicating their relationship to other continents of the Southern Hemisphere. (Courtesy of Cambridge University Press.)

of their biota and environmental characteristics overlap, ecosystems of the three zones are distinctively different.

1. Sub-Antarctic

The Sub-Antarctic zone consists of a ring of isolated sub-Antarctic islands and island groups lying at high latitudes in the Southern Ocean. With the exceptions of South Georgia, Heard Island, and McDonald Island, these lie either close to or north of the "Polar Frontal Zone" (formerly known as the Antarctic Convergence), a circumpolar oceanic feature formed where cold Antarctic surface waters sink below warmer sub-Antarctic waters. Sub-Antarctic island climates are strongly oceanic and, with the occasional exception of South Georgia, they are not influenced by seasonal pack or fast ice. Temperature variation is reduced by the surrouding cold ocean, with mean air temperatures of most being low but positive year-round, and precipitation is high (Table I).

Some islands remain extensively glacierized (South Georgia and Heard Islands) and formed the center of much larger ice caps extending well away from current coastlines at the height of Pleistocene glaciation. However, the majority are now or have been only partially glaciated. The age of these islands and availability of terrestrial habitats vary widely. South Georgia is composed of continental crust of Upper Jurassic/Lower Cretaceous age, once attached to the southern Andes. Kerguelen may be up to 100 million years old, but has never had a continental association, while Macquarie Island is formed from a segment of oceanic crust raised above the surface 3–5 million years ago. The remaining groups of Crozet, Prince Edward, and Marion Islands are younger, the latter apparently of fairly recent origin (ca. 300,000 years) and the former of Miocene origin.

Some authors also include more northerly island groups (Falkland Islands, Gough Island, Îles Amsterdam and St. Paul, and New Zealand's outlying groups

TABLE I

Typical Air and Microhabitat Temperature Ranges Experienced in Summer and Winter in Each of the Antarctic Biogeographical Zones[a]

		Air temperature range (°C)		Microhabitat temperature range (°C) (exposed habitat minima in parentheses)	
Antarctic zone	Months with positive mean air temperatures	Summer	Winter	Summer	Winter
Sub-Antarctic	6 to 12	−2 to +25	−10 to +15	−5 to +30	>−2 (−10)
Maritime Antarctic	1 to 4	−10 to +15	−45 to +5	−5 to +30	−20 to +3 (−45)
Continental Antarctic					
Coastal	0 to 1	−20 to +10	−40 to −5	<−10 to +30	(−37)
Inland (deserts and nunataks)	0	−31 to +9	−48 to −12	−29 to +17	−47 to −16

[a] Few data are available from microhabitats in the continental Antarctic, particularly during winter; considerably more negative minima would be expected in sites without snow cover.

of Snares, Campbell, Bounty, and Chatham) either within the sub-Antarctic or in a separate "mild Antarctic" classification. More correctly these should be classified as southern cool temperate or ocean temperate, containing faunal and floral groups not otherwise represented in the sub-Antarctic.

2. Maritime Antarctic

The maritime Antarctic also experiences significant oceanic influence on climate, but on a more seasonal basis than in the sub-Antarctic. The zone includes the western coastal regions of the Antarctic Peninsula to Alexander Island (ca. 72°S), the Scotia Arc island groups of the South Shetland, South Orkney, and South Sandwich Islands, and the isolated islands of Bouvetøya and Peter I Øy (see Fig. 1). Although in a biological sense the South Sandwich archipelago and Bouvetøya are within the maritime Antarctic zone, they are unusual in being geologically recent (1–3 million years old) volcanic islands. Most show evidence of recent or current volcanic activity, with unique biological communities associated with limited areas of geothermally warmed ground. Together with analogous areas on Deception Island (South Shetland Islands) and Mounts Erebus, Melbourne, and Rittmann (Victoria Land, continental Antarctic), such geothermally associated biota are exceptional within Antarctica.

Maritime Antarctic mean air temperatures are marginally positive for 1–4 months of the year; however, both summer maxima and winter minima are buffered to some extent by the surrounding ocean, the surface temperature of which varies annually between ca. −2 and +1°C (see Table I). Short periods of positive temperature and thaw are experienced in all winter months. Precipitation is generally high through a combination of oceanic influence and typical westerly airflow, but varies widely between sites and throughout the year.

3. Continental Antarctic

The continental Antarctic comprises most of the continental area of Antarctica, including all of East (or Greater) Antarctica, the Balleny Islands, and the eastern side of the Antarctic Peninsula. With the exception of the extensive ice-free cold desert region of the McMurdo Dry Valleys, terrestrial habitats of this zone are of limited extent and great isolation. In addition to the cold deserts, they include two basic habitat classes—exposed coastal regions superficially similar to those of the maritime Antarctic and inland nunataks. Temperature regimes are more extreme than those of the maritime Antarctic, with positive mean air temperatures achieved for <1 month in coastal locations and never inland. Air temperatures rarely if ever become positive even for short periods and, unlike in the maritime Antarctic, never in winter.

As in the other zones, microhabitat temperatures do not track air temperatures closely, the two most significant influences being warming due to absorption of solar radiation in summer and protection from winter extremes by snow cover. However, many continental Antarctic habitats also experience a period of prolonged darkness with no thermal input. Few data are available, but those habitats with no winter protection (e.g., exposed mountain ridges and cold deserts) are expected to track closely the winter air temperature patterns and minima.

II. ANTARCTIC TERRESTRIAL BIOTA

A. Habitats

In the maritime and continental zones, with the exceptions of steep cliffs and exposed mountain ridges, most terrestrial habitats are covered seasonally by snow and/or ice, which has the benefit of buffering them from extreme temperature lows and fluctuation, and also from wind abrasion. There is an important difference between the sub-Antarctic and other zones. While habitats in the latter zones may be free of seasonal snow cover for periods of only weeks up to perhaps 5 months (those in the most extreme continental sites may not be exposed at all in some seasons), many sub-Antarctic islands experience intermittent or no periods of snow cover, which is often restricted to higher altitudes. There is much variation, but even on the coldest islands such as South Georgia (with up to 6 months of annual snow cover in lowland sites), mean air temperatures do not drop far below zero, and sub-nivean microhabitat temperatures are often sufficient to allow year-round invertebrate activity.

Overwinter microhabitat temperatures in maritime and continental zones are normally low enough to arrest physiological processes and activity. For a period in the austral spring, absorption of solar energy by soils or plant communities may be sufficient to allow the formation of a sub-nivean air space. This acts almost as a "greenhouse," within which temperatures are high enough to allow positive levels of net photosynthesis by microbiota, lichens, and mosses and feeding activity by invertebrates (Convey, 1996; B. Schroeter *et al.*, in Lyons *et al.*, 1997). This feature effectively increases the active season length for these groups.

Antarctic soils are generally poorly developed, with low organic content. Again, there is a clear dichotomy between soils of the sub-Antarctic, and soils in the other two zones, with only the latter possessing a widespread and permanent permafrost layer. Brown soils, resembling those in temperate regions, are widespread in the sub-Antarctic zone but are associated only with the larger stands of higher plants (phanerogams) in the maritime Antarctic; they are not present in the continental zone. Likewise, deep peat deposits have developed under extensive valley bog communities in the sub-Antarctic, and have been radio-carbon-dated as forming soon after the end of the Pleistocene glaciation (10,000–20,000 years b.p.). Moss peat deposits are much more restricted in the maritime zone (5000–6000 years b.p.) and are not found in the continental Antarctic. Particularly in the maritime and continental zones, the formation, development, and stability of soils are heavily influenced by cryoturbation processes, leading to large areas of patterned ground (including soil/stone polygons, stripes, and solifluction lobes).

B. Terrestrial Fauna

Terrestrial vertebrates form a very small element of the natural fauna of Antarctica, and these few are limited to sub-Antarctic islands. They include a single endemic insectivorous passerine and freshwater ducks on South Georgia and/or Kerguelen. There are many records of vagrant birds, particularly from sub-Antarctic islands but also from the maritime Antarctic. The majority of these have obviously been carried far off course and succumb rapidly to the extreme conditions or predators. There are no naturally occurring mammals, reptiles, or amphibians, although human contact has led to the deliberate or accidental introduction and naturalization of a variety of mammals (see the following).

In contrast, marine vertebrates are plentiful in Antarctica, and penguins, seabirds, and seals spend significant periods ashore to breed, rest, and molt, often in colonies comprising tens of thousands or even millions of individuals. Most colonies are coastal, but birds such as the Antarctic and snow petrels and south polar skuas also breed on nunataks up to several hundred kilometers inland. These concentrations of marine vertebrates have direct influences on local terrestrial ecosystems through trampling and deposition of guano and carcasses. Aerosol dispersal of chemical nutrients from guano extends the influence of these colonies far beyond their physical boundaries.

In the absence of terrestrial vertebrates, the faunas of each Antarctic zone consist mainly of invertebrate groups (Table II). Those of sub-Antarctic islands include representatives of "higher" pterygote (winged) insects, although many of these show behavioral or morphological (brachyptery and aptery) traits that remove the ability to fly—a feature necessary to reduce the risk of being dispersed away from favorable habitats by frequent high winds. Many invertebrate groups well known in temperate areas are not found in the sub-Antarctic (e.g., Odonata and Trichoptera) or are represented by very few species (e.g., Isopoda, Araneae, Lepidoptera, Hymenoptera, and Hemiptera). The most diverse higher insect groups are Diptera and Coleoptera. Molluscs and annelid worms (Enchytraeidae) are also present in sub-Antarctic ecosystems, but there are very few records and no detailed studies available. Diverse communities of micro-arthropods (Acari and Collembola) and micro-invertebrates (Nematoda, Tardigrada,

TABLE II

Dominant Biotic Components of Typical Ecosystems of Each of the Antarctic Biogeographical Zones

Zone	Flora	Fauna	Microbiota
Sub-Antarctic	Phanerogams and cryptogams; open fellfield (cf. maritime zone) at higher altitude; alien species	Arthropods (esp. Insecta); microarthropods (Acari, Collembola); microinvertebrates (Nematoda, Tardigrada, Rotifera); alien vertebrates and invertebrates	Various groups
Maritime Antarctic	Closed and open cryptogamic communities; phanerogams very limited	Microarthropods (Acari, Collembola); microinvertebrates (Nematoda, Tardigrada, Rotifera); Diptera very limited	Algal and cyanobacterial mats; foliose algae; protozoa
Continental Antarctic	Bryophytes and lichens, of very limited extent	Microinvertebrates (esp. Nematoda, Tardigrada); microarthropods (Acari, Collembola) more limited	Algal and cyanobacterial mats; protozoa; endolithic fungi, algae, and cyanobacteria

and Rotifera) are also present in this zone, but have generally received less critical attention than the larger and more obvious groups.

Maritime Antarctic faunas include fewer taxonomic groups. The only higher insects represented are two chironomid midges (Diptera). The most-studied members of ecosystems in this zone are free-living soil microarthropods (Acari and Collembola), although micro-invertebrate groups (Nematoda, Tardigrada, and Rotifera) are also well represented numerically. All of these groups are known by a small number of species (see the following section), but population densities are often high, in the range 10^4–10^7 individuals m^{-2}, comparable with or greater than such densities in many temperate ecosystems. The arthropod fauna is subject to very low predation pressure, with a maximum of only two predatory mite genera (including three species) present in any ecosystem. Food webs therefore have a very simple structure. Microinvertebrate groups include species from genera with characteristic trophic preferences (e.g., algivory, bacterivory, fungivory, and nematophagy), however, the specific dietary preferences of Antarctic taxa are virtually unknown.

The terrestrial fauna of the continental Antarctic represents a further simplification of that found in the maritime zone. No higher insects are present, and micro-arthropods only reach the dominance and abundance seen in the maritime zone in the much more limited areas of vegetation and (vertebrate) nutrient enrichment. Instead, in the continental zone microfauna such as nematodes assume dominance. In the most extreme continental cold deserts, simple food webs consist of as few as 1–3 nematode species, only one of which may be predatory (Freckman and Virginia, 1997). These cold desert soils are faunistically the most sterile on Earth.

C. Terrestrial Flora

Changes in floral composition and patterns of diversity between the three Antarctic biogeographical zones mirror those of the faunal communities (see Table II). Smith (in Laws, 1984) provided the most authoritative description and comparative treatment of Antarctic floras, and it remains valid despite many subsequent advances in detailed taxonomic knowledge.

The flora of the sub-Antarctic is richest and most diverse, with well-developed phanerogamic and cryptogamic floras. These show strong affinities with either South American (Fuegian) or Australasian (New Zealand) floras. Trees and shrubs are not present (with the exception of the diminutive dwarf shrub *Coprosina* on Macquarie Island) and attempts to introduce several species to sub-Antarctic islands have been unsuccessful.

Sub-Antarctic vegetation can be described as "tundra-like." Although the vegetation is superficially similar, and climate, land-form, and pedology are comparable to those of Arctic tundra, the two regions differ in several important respects. Sub-Antarctic soils do not contain a permafrost layer, and the only woody plants are dwarf-shrub-like rhizomatous herbs (*Acaena*). The vegetation is dominated by large rosette-forming herbs ("megaherbs") and tall grasses, and includes a significant proportion of ferns. Phanerogamic vegetation is best developed in low-altitude coastal regions. As conditions worsen rapidly with increasing altitude, these plants are progressively replaced by a cryptogamic fellfield flora that is closely similar (and sharing many species) to that found in the more extreme maritime Antarctic.

The most favorable habitats of the maritime Antarctic are dominated by closed cryptogamic communities of carpet- and turf-forming mosses. Development of extensive vegetation is limited to a narrow altitude range (up to ca. 150 m) in coastal regions. Beyond this range, in inland areas, at higher altitudes, and at more exposed coastal sites, more open fellfield communities occur, consisting of several sub-formations of cushion- and turf-forming mosses, liverworts, and crustose, foliose, and fruticose lichens. Phanerogams are represented by only two species (hairgrass, *Deschampsia antarctica,* and pearlwort, *Colobanthus quitensis*), which are also found in the sub-Antarctic and across a wide latitudinal range in the South American Andes. Both are particularly sensitive to current climate amelioration in the Antarctic Peninsula region (see the following) and are undergoing rapid population increases.

Cryptogam-based maritime Antarctic vegetation is fragile and, lacking roots, sensitive to physical disturbance, for example, in the vicinity of penguin and seal colonies. Such areas tend to be dominated by the foliose alga *Prasiola crispa.* One consequence of the rapid recovery of the Antarctic fur seal (*Arctocephalus gazella*) populations in this century, following their near extinction through hunting and similar drastic human-mediated reductions in Southern Ocean cetacean populations (with which the seals are thought to compete for food), has been the occupation of new summer resting and molting sites within terrestrial habitats of the maritime Antarctic. The cryptogamic communities of these areas have been unable to withstand the twin pressures of excessive seal trampling and nutrient enrichment and have largely been destroyed in many coastal lowlands.

Continental Antarctic vegetation is categorized within the same system as that of the maritime zone, although lichen sub-formations predominate. No phanerogams are present, the extent of closed cryptogamic communities is much more limited, and peat formations are absent.

D. Microbial Systems

Microbial autotrophs form the basis of polar terrestrial ecosystem processes (Vincent, 1988; Friedmann, 1993; Wynn-Williams, 1996), playing pivotal roles in the processes of primary colonization and stabilization of mineral soils, which allow secondary colonization and succession by other microbiota, plants, and Metazoa. The most significant groups involved in primary colonization are autotrophic cyanobacteria and algae, which enable secondary invasion by bacteria, fungi, and protozoa (a polyphyletic group whose most significant members in Antarctica are heterotrophic flagellates, gymnamoebae, testate amoebae, and ciliates). At the level of microbial ecosystems, there is not a sharp divide between “terrestrial” and “aquatic” (stream/lake) habitats.

In addition to the largely edaphic (on or within soils), epiphytic (on surfaces of living plants and lichens), and epilithic/hypolithic (on exposed surfaces or undersurfaces of rocks) habitats typically occupied by faunal and floral communities, microbial ecosystems may also be found in cryophilic (between ice crystals in melting snow) and endolithic (within the surface few millimeters of rock matrix) habitats. The latter habitat is further divided into chasmoendolithic (within fissures and cracks open to the rock surface) and cryptoendolithic (within tiny cavities of the rock matrix). These cryptic habitats, as found in the cold deserts of continental Antarctica, are thought to represent one limit to biological existence on Earth, and as such they have been proposed as models to assist the development of exobiological assessment techniques (Wynn-Williams, 1996).

Microbial ecosystems present a visibly greater dominance in more extreme terrestrial habitats, although the same groups are present and important in all three Antarctic zones. Groups such as algae and cyanobacteria may become visible via the formation of filaments and mats, both within water bodies, such as streams and lakes, and on/in the surface layers of damp soils. Such communities are well represented within the maritime zone, and are often a climax community of the continental zone (see Table II). In the most extreme cold desert habitats of the latter zone, no detectable life survives on the surface of soils or rock. In these habitats, microbiota (algae, cyanobacteria, and fungi) have retreated to the endolithic niche, with cells existing in the interstitial spaces between rock crystals.

E. Freshwater Systems

Freshwater ecosystems of the Antarctic possess a very simple structure when compared with those of the Arctic and lower latitudes. Perhaps most striking is the absence of fish. The faunas of sub-Antarctic fresh waters include a single predatory diving beetle (found only on South Georgia), several Crustacea (cladocerans, copepods, ostracods, and anostracans), adventitious microarthropods, nematodes, tardigrades, rotifers, and protozoans. Ecosystem structure in the maritime Antarctic is even more simplified: the plankton include one copepod (*Boeckella poppei*) and one anostracan (*Branchinecta gaini*) herbivore, the predatory copepod *Parabroteas sarsi,* benthic cladocerans and ostracods, and the microscopic groups. Two Crustacea (*Daphniopsis studeri* and *Acanthocyclops mirnyi*) are recorded from the continental zone.

The trophic impact of metazoan groups in lakes of all three Antarctic zones is currently thought to be minimal, although data are lacking and there is disagreement over the importance of metazoan predators in the sub-Antarctic lakes of South Georgia. With this exception, top-down grazing control is reduced and many ecosystems are thought to be driven by bottom-up forces. In particular, the "microbial loop" assumes great importance—a microbial food web consisting of phytoplankton, bacteria, and protozoans (Laybourn-Parry, in J. Lyons *et al.,* 1997).

Lakes in the maritime and continental zones are seasonally or permanently covered by ice. Some continental lakes are meromictic or hypersaline. Unlike terrestrial ecosystems, some lakes are ancient systems thought to be hundreds of thousands of years old. At the extreme, Lake Vostok, a 200-km-long, 500-m-deep lake beneath at least 3 km of the continental ice cap, provides the intriguing possibility of harboring "ancient" microorganisms and other groups that have been effectively isolated since the formation of a permanent ice barrier (Karl *et al.*, 1999).

III. PRESENT-DAY BIODIVERSITY

A. Within Antarctica

Attempts to catalog and compare the biodiversity of the three Antarctic biogeographical zones are hindered by two fundamental problems: a lack of adequate (or comparable) sampling coverage, and taxonomic uncertainty (particularly the likelihood of extensive synonymy), both of which apply in varying extent to all groups encountered. Currently, sufficient data do not exist to allow rigorous comparisons for any microbial groups, beyond general evidence for lower diversity in the continental versus maritime and sub-Antarctic zones. The taxonomy of prokaryotic microbiota is poorly documented anywhere in the world, and the recent steps to develop molecular phylogenies are hard to reconcile with earlier classification systems, although efforts are accelerating. Both in the Antarctic and worldwide, a small proportion of prokaryotes have been described, with virtually no knowledge of species distribution. Thus, although it is possible to conclude from molecular comparisons that "the majority of Antarctic prokaryotes diverged from their nearest known non-Antarctic relatives long before a stable ice-sheet developed in Antarctica," it is not possible to estimate the time of colonization of the continent (Franzmann, in Wynn-Williams, 1996).

The biodiversity of groups for which there are reasonable data in each of the Antarctic zones is summarized in Tables III and IV. These data illustrate a general trend of reducing diversity, and loss of specific faunal and floral groups, along a transect of increasing environmental extremes from the sub-Antarctic to the continental Antarctic. Although there have been several proposals of similar trends occurring within zones (particularly along the Antarctic Peninsula within the maritime Antarctic), the true picture appears to be more complex, with biodiversity being related to the presence of suitable microhabitats, which themselves become reduced in extent at higher latitudes.

B. Bipolar Comparison

Environmental conditions experienced at high northern polar latitudes may be compared broadly with those of

TABLE III

Biodiversity of Plant Taxa in the Three Antarctic Biogeographical Zones[a]

Zone	Flowering plants	Ferns and club mosses	Mosses	Liverworts	Lichens	Macrofungi
Sub-Antarctic	60 native, 50 persistent aliens	16	250	85	250	70
Maritime Antarctic	2 native, 1 persistent alien	0	100	25	250	30
Continental Antarctic	0	0	25	1	150	0

[a] The figures presented are approximate, as it is likely that (1) new species records will be obtained through more directed sampling, (2) a significant number of unrecognized synonymies are likely to exist, and (3) taxonomic knowledge of some Antarctic groups is incomplete.

TABLE IV

Biodiversity of Native Terrestrial Invertebrate Taxa in the Three Antarctic Biogeographical Zones[a]

Group	Sub-Antarctic	Maritime Antarctic	Continental Antarctic
Protozoa[b]	83		33
Rotifera[b]	ND[c]	ND	13
Tardigrada	>29	26	20
Nematoda[b]	22	40	10
Annelida (Oligochaeta)	23	3	0
Mollusca	5	0	0
Crustacea (terrestrial)	6	0	0
Crustacea (nonmarine but including meromictic lakes)	48	11	11
Insecta (total)	210	35	49
Collembola	39	10	10
Mallophaga	61	25	34
Diptera	44	2	0
Coleoptera	40	0	0
Arachnida (total)	167	36	29
Araneida	20	0	0
Acarina	140	36	29
Myriapoda	3	0	0

[a] Based on W. Block, in Laws (1984), Pugh (1993, and pers. comm.), S. J. McInnes (pers. comm.), H. J. G. Dartnall (pers. comm.), and P. Convey (unpubl. data).

[b] Large changes are likely with future research due to current lack of sampling coverage, expertise, and/or synonymy.

[c] ND, number of representatives of group unknown.

the sub-Antarctic and maritime Antarctic zones, although with the exception of the Greenland ice cap, no Arctic climatic equivalence with the continental Antarctic is possible. Terrestrial habitats of the Arctic consist of the northern fringes of large continental landmasses surrounding a shallow ocean, whereas the Antarctic is itself a continental mass surrounded and isolated by a large extent of cold ocean. This simple geographical difference is sufficient to drive the large climate differences found between the two polar regions at any specific latitude.

Geography also leads to a fundamental difference in colonization patterns between the two regions. With continuous southwards land connection to North America, Europe, and Asia, colonization of Arctic regions following post-Pleistocene deglaciation was a simple process. In contrast, the extreme isolation of terrestrial and freshwater habitats in all Antarctic zones from plausible refugia or other continental sources of colonists, combined with the likely obliteration of all coastal habitats during glaciation, led to much lower colonization rates. As a result, contemporary levels of diversity are one or more orders of magnitude lower in the Antarctic than the Arctic. This statement holds true for many faunal, floral, and microbial groups (e.g., Longton, 1988; Danks, 1990; Wynn-Williams, 1996; Fogg, 1998).

C. Origin and Antiquity of Biota

As mentioned earlier, low-lying coastal terrestrial habitats of the maritime and continental zones are likely to have been obliterated during glaciation. Even in the sub-Antarctic, where some islands experienced incomplete glaciation, many terrestrial ecosystems currently based around complex plant communities did not exist more than 10,000–20,000 years ago, as evidenced by glaciological models and maximum radiocarbon dates obtained from peat deposits.

Because a large proportion of extant biota in all three zones is found in low-lying coastal areas, it is unlikely that they could have survived glacial maxima in refugia *in situ* (or in higher-altitude nunatak sites where, even now, they are not represented). Rather, their presence must be a result of post-glacial colonization from refugia on southern oceanic islands and/or continental sources. Identification of such refugia is at an early stage, and is restricted by the general lack of detailed biogeographical data for most groups, including a lack of relevant information from potential source regions of Southern Hemisphere continents.

As recent colonists, there is little evidence of evolutionary adaptation by biota to the extreme environmental challenges posed by the Antarctic. Even though many studies demonstrate the ecophysiological abilities of Antarctic organisms (see Block, 1990), all features identified to date are plesiotypic for the taxonomic group concerned. Their possession appears to be a prerequisite for colonization and establishment rather than an evolutionary response to conditions experienced. However, at least among the Acari, Collembola, and Nematoda, species endemic to one or more of the Antarctic zones are present, indicating that evolution has occurred since colonization.

Not surprisingly then, the biota of Antarctic zones is most closely related to that of other southern continents. However, a number of examples of plant and microbial groups with bipolar or alpine distributions are known (see Longton, 1988; P. A. Broady, in Wynn-Williams, 1996). These are groups with known ability to disperse via aerial propagules. Gross similarities between soil faunal communities of both polar regions (the dominance of herbivorous/detritivorous Acari, Collembola, and other micro-invertebrates) simply represent a consequence of the generalist function of these groups in soil ecosystems worldwide.

Nunataks and inland ice-free areas in the continental Antarctic have, in contrast, provided evidence of the continued existence of a relict fauna predating Antarctic glaciation (Acari, via patterns of specific and generic endemism; Marshall and Pugh, 1996), and biological support for the sequence of events involved in the breakup of the Gondwanan supercontinent (Tardigrada; McInnes and Pugh, 1998).

IV. NATURAL COLONIZATION PROCESSES

A. Current Colonization

Aerobiological sampling programs carried out in Antarctica demonstrate the presence of viable colonizing propagules in the airspora, from both "local" (Antarctic) and distant sources. Propagule densities are several orders of magnitude lower than obtained in comparable temperate and tropical studies. Most such records refer to microbial or lower plant groups (bacteria, cyanobacteria, fungi, algae, and bryophytes) and lichens; there are no verified records of the capture of living invertebrates from the air column.

Direct observational evidence of continuing contemporary colonization, on both short- and long-distance scales, is given by the discovery of developing populations on previously barren ground, and of previously-unrecorded species in known sites. Particularly compelling evidence is provided by the colonization of geothermally -heated ground associated with the few, very isolated, volcanically active sites spread around Antarctica. These sites are recent, short-lived, and of limited area. Their bryophyte (and presumably algal and microbial) communities contain species not found elsewhere in the Antarctic, with postulated ancestral populations at much lower latitudes in South America or New Zealand.

Evidence for recent natural colonization events involving the Antarctic terrestrial fauna is mostly lacking or equivocal. The dominant groups (in terms of diversity) in all three zones are members of the soil fauna, including micro-arthropods (mites and springtails) and other micro-invertebrates (nematodes, tardigrades, rotifers, etc.). All of these are small, inconspicuous, and easily overlooked, with many Antarctic distributional records being the results of the specific collecting efforts of individual taxonomic experts. Thus, for any of these groups, the detail of coverage both within and between Antarctic zones is very patchy, and therefore confirmation of any record as a "new" colonization event is virtually impossible.

Several sites in the maritime Antarctic and in Victoria Land (continental Antarctic) have hosted individual studies since the 1960s that are detailed enough to provide baseline micro-arthropod biodiversity data. However, with the exception of two detailed studies of the maritime Antarctic South Sandwich Islands separated by 33 years, no sites have received sufficient subsequent attention to allow comparisons to be made. This isolated volcanic archipelago possesses an entirely colonist micro-arthropod fauna, with no endemic species. Comparison of the two studies of this fauna is illuminating, as both identified several taxa of sub-Antarctic (South Georgian) origin, but very few of these were in common. In particular, species of oribatid mite (*Edwardzetes* and *Austroppia* spp.) recorded in the earlier study and springtail (*Cryptopygus caecus*) in the latter were sufficiently widespread not to have been overlooked by either. This "crusoe" pattern of records is suggestive of frequent colonization events and short-term establishment, followed by a high probability of extinction.

B. Mechanisms of Colonization

Colonization mechanisms may be classified into (1) directed active movement, (2) assisted transport by other species or nonliving debris, and (3) passive transport by air or water currents.

No members of the invertebrate Antarctic terrestrial fauna show evidence of directed migratory behavior and the first option is discounted in explaining current or short-term future patterns of Antarctic biodiversity. However, several vertebrate and invertebrate species with well-known active dispersal characteristics are being increasingly recorded as vagrants at sub-Antarctic and even maritime Antarctic locations, where arriving individuals may survive for long periods. Indeed, several Diptera and Lepidoptera have become established on sub-Antarctic Marion Island in recent years, and it is reasonable to expect other dispersing species to follow.

Assisted transport of non-parasitic invertebrates, plant seeds, and other propagules by vertebrates (birds), other invertebrates, and debris (e.g., driftwood) has been proposed as an important mechanism of Antarctic colonization by several authors, but no direct evidence exists to support this route for groups other than microbes. Transport is most likely to be rapid and via bird species that spend time associated with land-masses during the austral winter, such as skuas, gulls, and sheathbills (thus coming into contact with terrestrial biota), rather than marine mammals (seals) and birds that spend the winter period wholly at sea (penguins, albatrosses, and smaller petrels).

Passive transport in either air or water currents is also an attractive option, but again with little supporting evidence available for some groups. Some arthropods, such as oribatid mites (Acari) and springtails (Collembola), show rafting behavior or high tolerance of submersion in sea-water, and have large numbers of littoral or marine taxa. Others (e.g., prostigmatid mites) show no such ability. Transport in the air column presents the twin challenges of low temperature and desiccation. These conditions are most likely to be survived by organisms with specific dispersal (lichen soredia and moss spores) and/or resistant stages (tardigrade "tuns," anhydrobiotic nematodes, rotifers, and protozoan cysts), and so air transport is unlikely to be a viable mechanism of fortuitous transport for other groups.

C. Dispersing Propagules

Investment in features favoring dispersal is not a characteristic of the "adversity-selected" life histories typical of Antarctic terrestrial plants and animals. Rather, features of these life-history strategies include factors such as extended life histories, low reproductive output, investment in survival features, and development of behavioral and morphological traits that reduce the chance of "accidental" dispersal. Nevertheless, all Antarctic organisms require the ability to disperse over a range of scales, from the millimeters to centimeters required to move between or colonize areas of favorable microhabitat, through the meters to kilometers required to support local colonization of ice-free ground within the Antarctic, to the hundreds or thousands of kilometers necessary to allow colonization from lower latitudes. The wide distribution of most Antarctic species in suitable habitats, particularly of the maritime and continental zones, combined with the recent age of such habitats, argues for the success of dispersal of these species, despite the paucity of observational data.

Various Antarctic microinvertebrates (e.g., protozoans, rotifers, tardigrades, and nematodes) have the potential to disperse in a resistant desiccated (anhydrobiotic) state, and then simply rehydrate and resume activity upon arrival in a favorable microhabitat. Likewise, algal and cyanobacterial mats and mosses survive long periods of unfavorable conditions in a desiccated state, and are subject to fragmentation and wind dispersal. Lichens, bryophytes, and fungi possess sexually- and asexually-produced dispersing propagules. All of these are present in aerobiological samples collected in Antarctica, with evidence of both local and distant sources.

D. Atmospheric/Meteorological Circulation Patterns

The oceanic currents and typical pattern of strong westerly airflows encircling the Antarctic continent between approximately latitudes 40° and 60°S normally provide effective barriers to north–south movement of colonizing organisms or propagules by extending likely transit times to such an extent that survival may be impossible. There are several records of South American *Nothofagus* trunks stranded on beaches on the Antarctic Peninsula and maritime Antarctic archipelagos, but these may have circumnavigated the Southern Ocean before deposition. However, rapid transport of pollen of South American origin into the maritime Antarctic does occur, and is associated with large cyclonic air masses that infrequently create a north–south airflow. Transit time is then 1–2 days, with synoptic weather records suggesting that such events occur approximately once every 18 months. There is also clear evidence, from analysis of ice and lake sediment cores covering a long timescale, of deposition events involving volcanic ash or biological particles that originate from identifiable sources at lower latitudes in the Southern Hemisphere.

V. FUTURE TRENDS

A. Anthropogenic Influences

Both the continent and surrounding Southern Ocean were immune from human impact until recent times. Since the early nineteenth century, however, marine ecosystems have been devastated, at first by the uncontrolled exploitation of marine mammal (seal and whale) populations and, more recently, by the continuing impact of various fisheries either directly on their target species or on non-target bycatch. There is no reason to expect these ecosystems to return to their pre-disturbance states.

Human impact on the terrestrial environment commenced with visits of sealing and whaling vessels to sub-Antarctic and maritime Antarctic islands, followed by the first landings on the Antarctic Peninsula and continent in the late nineteenth and early twentieth centuries. Since the Second World War, many scientific research stations have been erected around the Antarctic, which brought the development of their logistic support operations and the importation of personnel and materials into the biome. This has been followed since about 1970 by rapidly increasing ship- and air-supported tourist operations. Antarctic terrestrial ecosystems are fragile and sensitive to disturbance, often existing on the same areas of ice-free ground where research stations are established and where large concentrations of wildlife attract tourist operations. Therefore, human presence has inevitably disturbed and destroyed local areas of terrestrial habitat.

Even before the expansion of direct impacts, the consequences of human activity were detectable throughout Antarctica (including the nearshore marine environment) in the form of chemical pollutants, whose increase is recorded throughout the industrial era, and radioactive decay products arising from the use or atmospheric testing of nuclear weapons. Both are dispersed to the continent by the same atmospheric circulation patterns as proposed for biotic dispersal.

1. Introduction of Alien Organisms

The most obvious direct biological impact of human activities occurs via the introduction of alien organisms. Accidental introductions and deliberate transplant experiments have shown that a wide range of flora, fauna, and microbes are capable of surviving and establishing viable populations in all Antarctic zones (Friedmann, 1993; Pugh, 1994). So far, the greatest impact has been on sub-Antarctic ecosystems, which have the longest record of human influence and the least extreme climate. Here, various vertebrates have been introduced both accidentally (rats and mice) and deliberately (fish, chickens, rabbits, cats, pigs, sheep, moufflon, cattle, and reindeer) to all major islands. In some cases, the consequences of these introductions are irreversible, as endemic species have been eradicated, while in others (particularly the presence of rats on larger islands) effective control measures appear to be impracticable. The ecological impact of some introductions appears negligible (e.g., trout to Kerguelen, mice to South Georgia) but, in general, appropriate studies have not been made. Introduced species may have both direct (e.g., predation of bird eggs and terrestrial invertebrates, trampling and grazing of plants) and indirect impacts (e.g., alteration of habitat structure leading to changes in species dominance or behavior) on native species.

Introductions of invertebrates are less well documented. There have been no deliberate introductions of invertebrate species to Antarctic sites. Rather, all such aliens have been introduced accidentally, with stores, food, equipment, or domestic animals associated with human settlement. Pugh (1994) estimated that 13.5% of ca. 520 Acari species reported from Antarctica (mainly the sub-Antarctic) originated from other continents, the majority with human assistance. Although many alien species have persisted only as long as human settlement continued, a number have become established (particularly in the sub-Antarctic, but also in the maritime Antarctic), with some evidence of competitive displacement of native species and the introduction of new trophic interactions to terrestrial ecosystems.

Introductions of flowering plants, bryophytes, and microbes have also accompanied human occupation. Again, the sub-Antarctic zone has been most affected—for instance, South Georgia now has more persistent alien than native flowering plants within its flora (25 versus 17 species, although many of the aliens have very restricted distribution and minimal impact). Transplant experiments have demonstrated that several sub-Antarctic and temperate vascular plants are capable of long-term survival in maritime Antarctic conditions, with one example of an accidental introduction (*Poa pratensis*) surviving for at least 30 years on the Antarctic Peninsula. Very few data on microbial introductions exist, although the process is inevitable via human transport (see Broady, in Wynn-Williams, 1996) and complete control measures are impracticable. Given the apparent evolutionary isolation of Antarctic prokaryotes (see earlier discussion), the potential for damage to this unique biological resource should not be underestimated.

2. Regional and Global Environmental Change

The identification and prediction of global climate change via global circulation models (GCMs) currently receive a great deal of attention from scientists, politicians, and the general public. Although details differ, most GCMs predict that any climate (temperature) amelioration will be both greatest and most rapid at high latitudes.

a. Temperature

There is clear evidence of rapid regional warming trends from long-term temperature records maintained at several maritime and sub-Antarctic research stations, with increases on the order of 1–1.5°C between 1950 and 2000. It is not yet possible to state conclusively whether this represents a regional or global process. There is limited evidence available from a small number of coastal sites in the continental Antarctic of a parallel process occurring, although it is not yet clear how any global processes will affect the main continental ice mass.

b. Water

Another important prediction of GCMs is changing patterns of precipitation, which alter the water input to terrestrial habitats. Detailed predictions are not available for Antarctica, although it is recognized that water is possibly the single most important factor limiting distribution of Antarctic terrestrial biota. Water availability from precipitation is modulated by seasonal temperature variation. At a local scale, water availability to terrestrial ecosystems can undergo drastic change as a result of factors as simple as the complete exhaustion of a snow bank or, conversely, the increased release of water from melting ice fields. These effects may be caused by changes in either the timing or amount of snow accumulation, or the duration of positive summer temperatures (i.e., in combination with temperature amelioration).

At a larger scale, there is abundant evidence of rapid glacial retreat in the sub-Antarctic and maritime Antarctic zones. If this process includes the increased rate or duration of melting during summer, it will clearly influence the rate of water input to terrestrial ecosystems. Conversely, a combination of more rapid melting of winter snowfall (a finite resource) and extended summer length may lead to *decreased* water availability as the summer progresses.

c. Ultra-violet Radiation

A recent anthropogenically generated influence on Antarctic and possibly global ecosystems—the ozone hole—has existed since the early 1980s. This depletion of the Earth's protective stratospheric ozone layer is caused annually in the austral spring over Antarctica by a light-catalyzed reaction between ozone and pollutants concentrated at high altitudes and latitudes by atmospheric circulation patterns. Its formation leads to a drastic reduction in the ozone concentration over Antarctica and high southern latitudes, allowing a disproportionate increase in the exposure of terrestrial (and shallow marine) ecosystems to potentially damaging shorter-wavelength UV-B radiation. Radiation levels experienced at any site vary depending on the pattern of movement of the ozone hole and modulation by factors such as clouds and snow cover, but at worst exposure levels comparable to those of the tropics are experienced. This represents a very rapid change in a biologically important variable over a period of less than 20 years.

B. Biological Consequences of Climate Change

Despite a predictive literature (e.g., Voytek, 1990; Wynn-Williams, 1994) and increasing numbers of published studies of species in the Arctic region (e.g., Oechel *et al.*, 1997; Henry, 1997), identification of the effects of climate change processes in the Antarctic terrestrial environment is only now starting to receive critical attention (see Kennedy, 1995; Convey, 1997), and it is currently impossible to predict the detailed trajectory of change for any specific ecosystem. Knowledge of the physiological and life-history characteristics of individual species indicates that extant Antarctic terrestrial biota exhibit sufficient physiological and ecological scope or flexibility to absorb and even benefit from both the direct and indirect (e.g., changes in nutrient availability) effects of predicted levels of change. However, how individual species responses may be integrated at the community or ecosystem level is unknown in Antarctic systems (but see Freckman and Virginia, 1997). Any climate amelioration is likely to influence Antarctic ecosystems and biodiversity further by easing the constraints that limit colonization and establishment of exotic species.

The effects of temperature amelioration are expected to be greater at high-latitude sites, which already have a very restricted thermal energy budget and where the relative importance of a small temperature increment will be of greater significance. Thus an expected consequence of amelioration (assuming continued or increased water availability) is to increase the effective length of season and/or thermal energy budget of terrestrial ecosystems, which should in turn lead to increased growth, reproductive rates, and population sizes of both photosynthetic autotrophs and heterotrophic microbial and invertebrate species. The limited evidence available from Antarctic studies supports this contention, with increasing growth rates and more rapid completion of development reported in experimental studies of higher plants and invertebrates, combined with field observations of rapid areal increases in plant communities (Convey, 1997). Parallel studies from the Arctic are more equivocal, where groups such as aphids showed large population increases (11-fold increase in numbers of overwintering eggs) in response to realistic temperature manipulations, while other detritivorous soil arthropod populations did not show a detectable response (Hodkinson *et al.*, 1998).

Some climate change processes are likely to have deleterious effects on Antarctic biota. In particular, any decrease or total loss of water input to ecosystems could lead to local extinctions and drastic changes in ecosystem structure (via both physical and biological effects). Likewise, should the maximum exposure to increased UV-B radiation surpass the tolerances of exposed biota, particularly colonizing microbiota with crucial roles in soil stabilization that permit secondary colonization by other organisms, some terrestrial habitats may become barren.

Because of the extreme isolation of terrestrial habitats in all three biogeographical zones from potential source populations in the lower latitudes, initial responses to climate change processes will be restricted to the extant biota. However, climate amelioration will lead to both an increased area of available terrestrial habitat and a longer "window of opportunity" within which to complete the dual processes of colonization and establishment. Eventually, this would likely lead to increases in biodiversity and/or shifts in community composition, with or without human assistance, and in turn to greater trophic complexity and the inclusion of higher trophic levels. The resistance of extant Antarctic biota to these processes is unknown. Yet these simply structured ecosystems do present ideal conditions with which to test important ecological theories relating to community stability and species redundancy.

See Also the Following Articles

ARCTIC ECOSYSTEMS • AUSTRALIA, ECOSYSTEMS OF • DISPERSAL BIOGEOGRAPHY • SOUTH AMERICA, ECOSYSTEMS OF

Acknowledgments

I thank the British Antarctic Survey (BAS) for resources, many BAS staff for discussions that have contributed to the development of this paper, and two anonymous referees. Dr. H. J. Peat extracted biogeographical information contained in Table II from databases held by the BAS Biological Sciences Data and Resources Centre.

Bibliography

Block, W. (1990). Cold tolerance of insects and other arthropods. *Philos. Trans. Roy. Soc. B* **326**, 613–633.

Convey, P. (1996). The influence of environmental characteristics on life history attributes of Antarctic terrestrial biota. *Biol. Rev.* **71**, 191–225.

Convey, P. (1997). Environmental change: Possible consequences for the life histories of Antarctic terrestrial biota. *Korean J. Polar Res.* **8**, 127–144.

Danks, H. V. (1990). Arctic insects: Instructive diversity. In *Canada's Missing Dimension: Science and History in the Canadian Arctic Islands* (C. R. Harington, ed.). Canadian Museum of Nature, Ottawa.

Fogg, G. E. (1998). *The Biology of Polar Habitats.* Oxford University Press, Oxford, United Kingdom.

Freckman, D. W., and Virginia, R. A. (1997). Low-diversity Antarctic soil nematode communities: Distribution and response to disturbance. *Ecology* **78**, 363–369.

Friedmann, E. I. (1993). *Antarctic Microbiology.* John Wiley & Sons, New York.

Henry, G. H. R. (ed.). (1997). The international tundra experiment (ITEX). Short-term responses of tundra plants to experimental warming. *Global Change Biol.* **3**, supplement 1.

Hodkinson, I. D., Webb, N. R., Bale, J. S., Block, W., Coulson, S. J., and Strathdee, A. T. (1998). Global change and Arctic ecosystems: Conclusions and predictions from experiments with terrestrial invertebrates on Spitsbergen. *Arct. Alpine Res.* **30**, 306–313.

Karl, D. M., Bird, D. F., Björkman, K., Houlihan, T., Shackleford, R., and Tupas, L. (1999). Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science* **286**, 2144–2147.

Kennedy, A. D. (1995). Antarctic terrestrial ecosystem responses to global environmental change. *Annu. Rev. Ecol. Syst.* **26**, 683–704.

Laws, R. M. (ed.). (1984). *Antarctic Ecology,* Vol. 1. Academic Press, London.

Longton, R. E. (1988). *The Biology of Polar Bryophytes and Lichens.* Cambridge University Press, Cambridge, United Kingdom.

Lyons, W. B., Howard-Williams, C., and Hawes, I. (eds.). (1997). *Ecosystem Processes in Antarctic Ice-Free Landscapes.* Balkema, Rotterdam, Netherlands.

Marshall, D. J., and Pugh, P. J. A. (1996). Origin of the inland Acari of continental Antarctica, with particular reference to Dronning Maud Land *Zool. J. Linn. Soc.* **118**, 101–118.

McInnes, S. J., and Pugh P. J. A. (1998). Biogeography of limno-terrestrial Tardigrada, with particular reference to the Antarctic fauna. *J. Biogeogr.* **25**, 31–36.

Oechel, W. C., Cook, A. C., Hastings, S. J., and Vourlitis, G. L. (1997). Effects of CO_2 and climate change on arctic ecosystems. In *Ecology of Arctic Environments* (S. J. Woodin and M. Marquiss, eds.), Chap. 12. Blackwell, Oxford, United Kingdom.

Pugh, P. J. A.(1993). A synonymic catalogue of the Acari from Antarctica, the sub-Antarctic islands and the Southern Ocean. *J. Nat. Hist.* **27**, 323–421.

Pugh, P. J. A. (1994). Non-indigenous Acari of Antarctica and the sub-Antarctic islands. *Zool. J. Linn. Soc.* **110**, 207–217.

Vincent, W. F. (1998). *Microbial Ecosystems of Antarctica.* Cambridge University Press, Cambridge, United Kingdom.

Voytek, M. A. (1990). Addressing the biological effects of decreased ozone on the Antarctic environment. *Ambio* **19**, 52–61.

Wynn-Williams, D. D. (1994). Potential effects of ultraviolet radiation on Antarctic primary terrestrial colonizers: Cyanobacteria, algae, and cryptogams. *Antarct. Res. Ser.* **62**, 243–257.

Wynn-Williams, D. D. (ed.). (1996). Antarctic microbial diversity. *Biodiversity and Conservation* **5**(11) (special issue).

AQUACULTURE

Nils Kautsky,* Carl Folke,* Patrik Rönnbäck,* Max Troell,*
Malcolm Beveridge,† and Jurgenne Primavera‡
*Stockholm University and Royal Swedish Academy of Sciences, †University of Stirling, and
‡Southeast Asian Fisheries Development Center

GLOSSARY

aquaculture The farming of aquatic organisms, including fish, mollusks, crustaceans, and aquatic plants. Farming implies some sort of intervention in the rearing process to enhance production, such as regular stocking, feeding, or protection from predators. Farming also implies individual or corporate ownership of the stock being cultivated. For statistical purposes, aquatic organisms which are harvested by an individual or corporate body which has owned them throughout their rearing period contribute to aquaculture, whereas aquatic organisms which are exploitable by the public as a common property resource, with or without appropriate licenses, are the harvest of fisheries.

broodstock Fish or shellfish from which a first or subsequent generation may be produced in captivity, whether for growing as aquaculture or for release to the wild for stock enhancement.

farming intensity In a broad continuum, extensive systems are those which are closest to natural fisheries, requiring minimal inputs and offering relatively low yields, whereas intensive systems require a large amount of inputs to maintain an artificial culture environment, with high yields. Between these extremes are the varying degrees of semi-intensive aquaculture, where definitions are less distinct: (i) extensive aquaculture does not involve feeding of the organism, (ii) semi-intensive aquaculture involves supplementation of natural food by fertilization and/or the use of feeds, and (iii) intensive aquaculture is when the culture species is maintained entirely by feeding with nutritionally complete diets.

feed conversion The efficiency of farmed animals to incorporate given feed into biomass. Feed conversion is usually expressed in terms of the feed conversion ratio of weight of diet used to fish/shellfish flesh biomass produced. The ratio is affected by the relative moisture content of both feed and aquaculture product as well as the metabolic characteristics of the farmed species, farming techniques, and husbandry.

seed A term used to describe eggs, larvae, postlarvae, or juveniles (fry and fingerlings) stocked into aquaculture production systems.

spawner Mature individual of a stock responsible for reproduction.

AQUACULTURE, the aquatic counterpart of agriculture, has grown rapidly in recent decades to become one of the most important means of obtaining food from the sea. Impacts of aquaculture on biodiversity arise from the consumption of resources, such as land (or space), water, seed, and feed, their transformation into products valued by society, and the subsequent release into the environment of wastes from uneaten food, fecal and urinary products, and chemotherapeutants as well as microorganisms, parasites, and feral animals. Negative effects may be direct, through release of eutrophicating substances, toxic chemicals, the transfer of diseases and parasites to wild stock, and the introduction of exotic and genetic material into the environment, or indirect through loss of habitat and niche space and changes in food webs. Today, large quantities of fish are caught to produce fish meal—the main ingredient in feed—which may result in overfishing and affect marine food chains, including marine mammals and top carnivores. In some types of aquaculture, fish and shrimp larvae are caught in the wild to be used as seed. This may also result in bycatches of large amounts of other larvae, representing losses to capture fisheries and biodiversity. Large areas of critical habitats such as wetlands and mangroves have been lost due to aquaculture siting and pollution, resulting in lowered biodiversity and recruitment to capture fisheries. The magnitude of biodiversity loss generally increases with scale, intensity of resource use, and net production of wastes, but it is very much dependent on which species is cultured and the method of cultivation. In some cases aquaculture may increase local biodiversity, e.g., when ponds are constructed in dry areas and with integrated aquaculture.

I. AQUACULTURE AND BIODIVERSITY

A. The Aquaculture Process

Aquaculture is an economic human activity that uses and transforms natural aquatic resources into commodities valued by society, e.g., fish, shrimp, mussels, and seaweed. In doing so it may impact on biodiversity, essentially due to the consumption of resources, the transformation process (aquaculture), and the production of wastes (Fig. 1).

Contrary to common belief, technical and economic inputs, such as construction materials, energy, and labor, form only a small part of the inputs needed for aquaculture. The main and critical inputs are instead natural resources. Together with nature's services they ultimately determine the limits for the local and global expansion of aquaculture. The magnitude and type of resource use and impacts of aquaculture are very much dependent on which species is cultured, the farming methods used, and the intensity of farming (Box 1 and Table 1).

The aquaculture process in itself may affect biodiversity as a result of disturbance through increased road and boat traffic. High densities of farmed fish and food often attract predators and scavengers such as wild fish, gulls, and seals, These can come into conflict with farmers and may be killed, either accidentally (entanglement

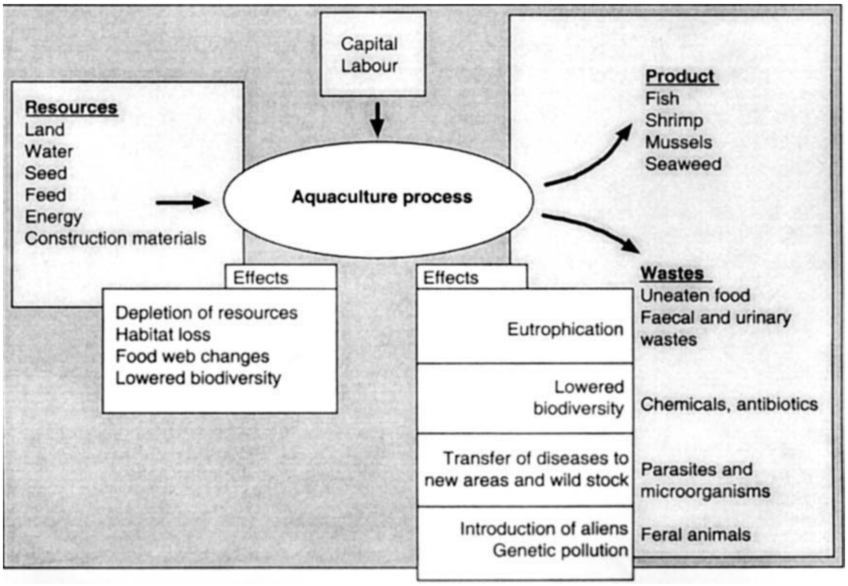


FIGURE 1 Diagram summarizing direct and indirect effects on biodiversity of the aquaculture process through the use of resources and the generation of wastes. Details are given in the text.

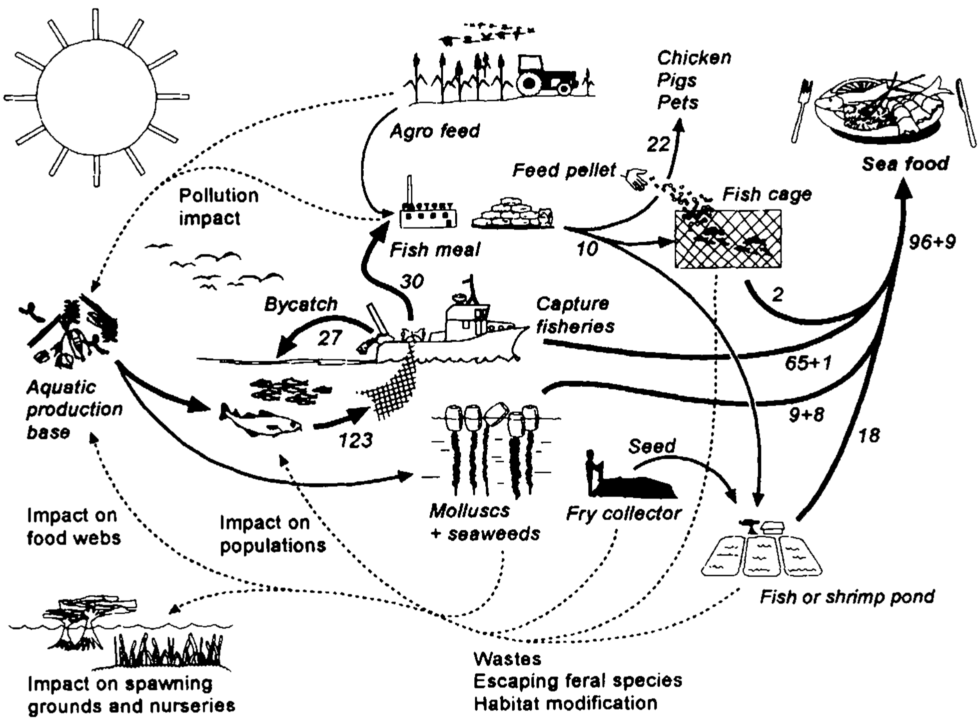


FIGURE 2 Ecological links between capture fisheries and various types of aquaculture: fish mollusk and seaweed rearing. Thick lines refer to main flows from aquatic production base via fisheries and aquaculture to human consumption of seafood. The thin line refers to agricultural input to aquaculture feed production. The numbers refer to 1997 million metric tons of fish. Dashed lines indicate negative feedbacks and possible impacts on biodiversity (reprinted from *Nature*, Naylor *et al.*, © 2000, Macmillan Magazines Ltd.). Details are given in the text.

in nets) or deliberately (shooting and trapping), or if they become established they may displace sensitive local species.

B. Feed

Whereas extensive aquaculture of herbivorous species such as carps and filter-feeding bivalves uses natural production and semi-intensive farming is supplied agricultural wastes and some feed, intensive aquaculture of fish and shrimp uses formulated feeds based on fish meal and fish oils (Boxes 1 and 2 and Table I). Many intensive and semi-intensive aquaculture systems use two to five times more protein, in the form of fish meal, to feed the farmed species than is ultimately harvested.

Feed requirements place a strain on wild fish stocks, and currently about one-third of the total harvest of capture fisheries is used to produce fish meal, one-third of which is used by the aquaculture industry. This may result in overfishing of small pelagic species, affecting marine food chains and ultimately marine mammals and top carnivores.

C. Land

Land is needed for building fish or shrimp ponds, whereas fish cages, pens, and mussel and seaweed farms occupy space in lakes and the sea. Globally, fish ponds are usually sited in agricultural land and this arguably contributes to landscape and floral and faunal diversity. In Europe, however, unproductive, boggy areas of agricultural land have often been used, and since such boundary ecosystems, or ecotones, may serve as reserves for species in areas otherwise surrounded by monocultures of crops, this may reduce biodiversity. Large areas of tropical coastal wetlands and mangroves have been converted to fish and shrimp ponds, resulting in lowered biodiversity and recruitment to fisheries. When the full range of ecological effects associated with mangrove habitat loss are accounted for, the net production in fish and shrimp aquaculture may be negative (Box 3).

It is increasingly being realized that the direct land and space resources occupied by agricultural or industrial production facilities are a side issue; what really

TABLE I

Resource Use in Relation to Intensity of Culture and Assessment of the Significance of the Sector

Farming intensity	Natural resource use	Relative importance
extensive		
seaweed	construction materials (ropes, wood); insignificant	highly significant; global; 7.2 mmt
molluscs (mussels, oysters)	construction materials (ropes, wood, bags/nets); insignificant	highly significant; global; 8.6 mmt
shrimp, crabs and lobster	land, (sea)water, seed (wild)	insignificant; tropics/sub-tropics; intensification widespread (<0.1 mmt)
fish (carps, tilapias, salmonids, etc.)	land, (rain)water, seed (usually wild capture)	insignificant; of local importance in tropics only; limited to poorer sectors in a few societies (<0.1 mmt)
semi-intensive		
shrimp, crabs and lobster	land (significant), seawater (insignificant), seed (wild; significant), feed (increasingly significant)	significant; tropical; ~0.8 mmt
fish	land (significant), water (insignificant), seed (usually wild; significant), feed (increasingly significant)	highly significant; tropics and sub-tropics; ~16.5 mmt
intensive		
shrimp, crabs and lobster	land (significant), seawater (insignificant), seed (wild; significant), feed (significant)	insignificant and decreasing; tropics and sub-tropics; ~0.3 mmt
fish	land (significant), water (significant), seed (hatchery and wild; moderately significant), feed (highly significant and increasing)	significant, especially in temperate and sub-tropics; increasingly important; ~2.5 mmt

matters is the area required to supply the resources that sustain the activity (see Box 6).

D. Water

Aquaculture requires large amounts of clean water to support the farmed animals, replenish oxygen, and remove wastes. In land-based systems, aquaculture does not only borrow water, returning it in a more degraded form, but also consumes it or accelerates its loss from surface to groundwater or the atmosphere. Thus, by creating ponds, especially in areas of poor (sandy/loam) soils or high temperatures, evaporation and seepage are increased and as much as 1–3% of the fish pond volume may be lost in this way each day. Such losses may be particularly significant in arid or semi-arid areas of the world, such as Israel, where fish pond design and management practices have had to be changed in order to reduce surface water losses. Conversely, the incorporation of a fish pond into small rural farms has been shown to improve water conservation.

E. Seed

Seed (eggs, larvae, and fry) is required to stock fish and shellfish farms. Although most inland and temperate coastal aquaculture relies on hatchery-produced fry and fingerlings, tropical marine fish and shrimp culture is still largely dependent on wild broodstock or juveniles, often creating demands that adversely affect wild stocks. Moreover, it is not only the target species which may be at risk since large bycatches of other larvae are killed, representing losses to capture fisheries and biodiversity (Box 4).

F. Impacts of Wastes

The term “wastes” in the current context is used to mean not only food, fecal and urinary products, and chemicals but also microorganisms, parasites, and feral animals which may be introduced to new environments along with aquaculture and may escape from aquaculture facilities. These have implications on biodiversity (Fig. 1 and Box 5). Uneaten food and fecal and urinary wastes may lead to eutrophication and oxygen depletion, the magnitude of which is dependent on the type and size of operation and the nature of the site, especially size, topography, and water exchange. Chemotherapeutants, including antimicrobial compounds and pesticides, are used in intensive marine fish culture to control bacterial and parasitic diseases. Controls on use are increasingly strict, especially in northern Europe and North America. In shrimp farming, many chemicals are used, largely without regulation. Farmers use a range of vitamins, immunostimulants, disinfectants, and chemotherapeutants and employ chemicals for pond soil and water treatment. The impacts of these chemicals are largely unknown, although there is evidence that microbial biodiversity can be affected by antibiotics, and antiparasitic chemicals may have implications for biodiversity.

G. Assessment of Resource Use and Carrying Capacity

To reduce the risk of resource constraints and impacts on biodiversity, a shift to aquaculture production systems that use less resources and emit wastes that do not exceed the assimilative capacity of the environment must occur. We also need to recognize and manage nature's life support on which economic development and human welfare depends. One way of identifying the demands for natural resource and ecosystem services of aquaculture is to estimate the ecosystem area—the ecological footprint—functionally required to support the activity (Box 6).

The footprint concept has proven to be very useful in illuminating the nonpriced and often unperceived work of nature that forms the basis for economic activities such as industrial aquaculture. When problems appear in fish ponds or fish cages, people tend to look at what is going on inside the pond or cage, not realizing that the farm is part of a much larger surrounding ecosystem, with which it interacts. Surrounding ecosystems provide the feed, seed, clean water, and other necessary natural resources and ecosystem services including waste assimilation. This work of nature sets the limits on how much can be cultured without compromising biodiversity or creating pollution or disease problems. Exceeding the carrying capacity of the environment is a major reason why many aquaculture developments, such as intensive shrimp farming, have had problems or even failed.

H. Improving Aquaculture Practices

The aquaculture industry has made considerable achievements in areas in which environmental impacts have negatively affected production quantity and quality. Some feeds have been formulated with lipids instead of proteins as an energy source in order to reduce nitrogen excretion by fish (Autin, 1997). Plant proteins with lower phosphorous levels are also used in feeds to reduce water pollution since most fish meal contains more phosphorous than fish can assimilate (Rumsey, 1993). In addition, the use of extrusion processes to develop high-quality fish meals has improved protein digestibility and produced more buoyant feeds that reduce waste (Botting, 1991), whereas the so-called "smart" feeding systems that monitor appetite and adjust feeding regime accordingly reduce food wastage considerably.

Many other systems have also been created in response to environmental damages and consumer concerns. In Orkney, Scotland, salmon produced with certified organic labels are now available for sale on the international market (Seafarm Business Review, 1998; Aarset *et al.*, 1998).

In some cases, the development of polyculture and integrated farming systems that make more efficient use of inputs and generate less waste, thus adding to net fish supplies locally and regionally at relatively minor environmental and social expense, may be a viable option (Troell *et al.*, 1997) (Box 7).

These examples indicate the positive directions in which the industry is heading. However, in cases in which awareness of environmental consequences is low, or in which the impacts do not affect profitability of the farmers and retailers, there is little incentive for the industry to regulate itself. Whereas economics and markets are promoting the adoption of technical innovation among different sectors of the industry, the policymakers must also take responsibility by incorporating proper incentive structures, developments of integrated coastal zone management, environmental impact assessment, and other organizational and institutional frameworks, including trade rules and regulations.

II. CONCLUSION

Since the principle of aquaculture is to reroute flows of energy and matter from the ecosystem into those species that we culture, aquaculture, like agriculture, will always affect the environment to some extent. This interaction and alteration of supporting environment is unavoidable, but it should not be done in a fashion that deteriorates the capacity of ecosystems to sustain social and economic development. Recent research has revealed that aquaculture systems developed in the past few decades are not an independent food-producing process. They are strongly coupled to nature's subsidies and services to sustain production.

Species such as shrimp, salmon, and other intensively reared carnivorous species in coastal areas present problems for wild fish stocks since feed requirements alone lead to a net loss in fish protein for these species

(Box 2). In addition, the conversion of mangroves and other coastal ecosystems, the collection of wild larvae for some species, and the spread of exotic species, diseases, and effluents create a spiral whereby aggregate fish supplies are reduced over time.

The farming of filter-feeding bivalves, herbivorous fish, and omnivores that are flexible in their dietary requirements places significantly less pressure on coastal and ocean resources. Most of the important species of farmed carp, tilapia, and mollusks, for example, are herbivores or omnivores, and they account for a large share of fish protein consumed in the developing world.

The real challenge for aquaculture is to develop farming practices that are in tune with ecosystem processes and functions in a fashion that enhances seafood production. There is great potential to develop techniques that work with nature's dynamics. There is also potential to redirect unsustainable modes of production into practices that contribute and enhance nature's support capacity not only for aquaculture but also for other human activities dependent on aquatic ecosystems. There is no doubt that the aquaculture business will move in this direction. Governmental policies and institutional frameworks are required that can make such a transition possible. The internalization of the costs of deteriorated coastal and marine support areas caused by the farming in the price of the product, i.e., applying the "polluter pays principle" in line with the Rio Declaration, would create incentives for the industry to take a more sustainable path (Folke *et al.*, 1994). This is not easy to accomplish, however, because it will have to be done simultaneously for all food production systems (i.e., fisheries, agriculture, and animal production) in order not to disrupt world markets.

The role of the consumer may also be critical in shaping farm management practices. Aquaculture products are increasingly traded through multinational supermarkets that are highly responsive to customer opinion and demands. If farmed aquatic foods become associated in the public's mind with poor environmental management and its direct effects on biodiversity, then supermarkets may well refuse to stock the produce.

Box 1

Aquaculture Development and Practices

Aquaculture, the farming of aquatic plants and animals, is many thousands of years old. Nevertheless, it must be regarded as a largely post-World War II phenomenon. In 1950, global farmed fish and shellfish production was approximately 2 million tons and largely confined to areas of Asia. During the past 50 years, global aquaculture production has grown nearly 20-fold to 36 million tons in 1997. In the 10-year period ending in 1997, production of farmed fish increased, resulting in a more than doubling in weight and value, as has its contribution to world fish supplies (FAO, 1999). The aquaculture sector today produces 16.7 million tons of finfish (49% of the total), 1.1 million tons of crustaceans (3%), 8.5 million tons of mollusks (25%), and 7.7 million tons of aquatic plants (23%). Fish produced from farming activities, or aquaculture, now accounts for more than one-fourth of all fish directly consumed by humans. More than 300 species of fish and shellfish are farmed; the range includes giant clams that obtain most of their nutrients from symbiotic algae, various species of carps that are largely herbivorous, and Atlantic salmon and marine fish species that are carnivorous (Williams, 1996). Aquaculture typically involves the enclosure of a species in a secure system under conditions in which it can thrive. Interventions in the life cycle range from exclusion of predators and control of competitors (extensive aquaculture) to enhancement of food supply (semi-intensive) or even the provision of all nutritional requirements (intensive). Intensification also implies increasing the number of individuals per unit area, which requires greater use and management of inputs and a greater reliance on technology and fossil energy.

Production practices and their impacts on aquatic ecosystems vary widely across species (Table I). Mollusks such as scallops and mussels are generally farmed along subtidal or intertidal coastlines where wild-caught or hatchery-reared seed are grown in bags set on the sea bottom or on stakes and suspended ropes. The animals rely entirely on prevailing supplies of plankton and organic particles for food. A range of systems—ponds, tanks, or cages—are used to farm finfish. The majority of carp and other freshwater species farmed in the tropics and sub-tropics are herbivores/omnivores and are grown in fertilized by supplemental feeds (see Box 2). In contrast, most diadromous and marine finfish, including both tropical and temperate species, are farmed intensively in floating net cages and are reliant on nutritionally complete fish meal and fish oil-based diets. Crustacean farming is dominated by penaeid shrimp. Shrimp postlarvae, either derived from captured wild parents or directly collected from the sea, are stocked in coastal ponds and grown for 3 or 4 months until they reach marketable size. The more intensive shrimp systems depend mainly on formulated pellet feeds, aeration to replenish dissolved oxygen, and pumped seawater to dilute pollutants and flush out harmful metabolites.

Box 2

Fisheries for Fish Meal as Feed

Extensive aquaculture, such as that of seaweed and mollusks, is solely reliant on ambient nutrient/food supplies. The traditional semi-intensively managed fish and shrimp production systems that predominate in the tropics and subtropics also use little or no fish meal. Herbivorous and omnivorous carps and tilapias are grown in ponds to which nutrient-rich materials are often added to stimulate growth of algae and other food items. On-farm feeds, based largely on cheap, locally available agricultural by-products such as rice bran, augmented by household scraps and perhaps small amounts of fish meal are used to supplement the food in the ponds. However, in the intensive production systems that predominate in temperate aquaculture the farmed animals are solely reliant on fish meal and fish oil-based diets.

Only 2 fish of the 10 most commonly farmed species, catfish and carp, require less fish as inputs than is ultimately harvested (Naylor *et al.*, 2000). For the others more fish is required as feed inputs than is ultimately produced: 2 kg of wild fish is used to produce 1 kg of farmed salmon, trout, and marine shrimp, and 4.5 kg of wild fish is used to produce 1 kg of marine finfish and eel (Tacon, 1996). The culturing of these species will thus lead to a net loss in fish protein.

Human consumption of seafood is 103 million metric tons (mmt), of which 65 mmt of fish, crustaceans, and mollusks and 1 mmt of seaweed come from capture fisheries, whereas 29 mmt of fish and 8 mmt of seaweed are from aquaculture (Fig. 2) (Tacon, 1998). Global fish capture is much higher—123 mmt—but 27 mmt of this is discarded as bycatch (Alverson *et al.*, 1994). and 30 mmt of it is used for fish meal production. An additional 2 mmt of processing scraps from aquaculture and fisheries is also converted into fish meal (Pike, 1998). Two-thirds of total fish meal supplies (22 mmt) are currently used in feed for chickens, pigs, and other animals, and one-third (approximately 10 mmt) is converted into aquaculture feeds (Tacon, 1998; Pike, 1998). However, an increasing proportion will probably be used for aquaculture feeds as supplies are unlikely to expand and as aquaculture production continues to grow and production methods of pond fishes in major producer countries such as China intensify.

Diets for salmonids, seabass, and sea bream and other carnivores are largely composed of fish meal and fish oil. Although it may be possible to replace much of the fishmeal used in intensive fish diets with oilseed proteins (Stickney *et al.*, 1996), requirements for essential amino acids, especially cystine and methionine, will continue to have to be met from other sources such as fish meal. It remains to be seen whether commercial oilseed protein-based diets can be developed in an industry in which the product is competing with many others for customer attention and in which profit margins are increasingly being squeezed. Depending on source and inclusion rate, oilseed meals can compromise palatability, growth (Stickney *et al.*, 1996) and, possibly, profitability. Any decrease in palatability or diet digestibility may aggravate waste loadings to the environment (Beveridge *et al.*, 1999). The issue of fish oils is perhaps even more pressing than that of fish meal. Aquatic carnivores are poor at using carbohydrate to supply energy requirements, relying instead on protein and lipid (Cowey and Sargent, 1977). The substitution of fish oils with vegetable oils in freshwater carnivorous or omnivorous fish diets is possible (Buzzi *et al.*, 1997). However, there are limitations with regard to marine carnivores such as salmon because they require n-3 highly unsaturated fatty acids which currently can only be derived in commercial quantities from fish oils (Bell *et al.*, 1986).

The use of fish meal and fish oil in livestock and aquaculture feeds reduces aquatic productivity through a series of food web interactions. Four of the top 5 and 8 of the top 20 capture species are used for reduction to fish meal (FAO, 1998). All are small, pelagic fish, including anchoveta, Chilean jack mackerel, Atlantic herring, chub mackerel, Japanese anchovy, round sardinella, Atlantic mackerel, and European anchovy. Many fishes exploited for feed are overfished and are strained by climatic variability associated with El Nino–Southern Oscillation events (FAO, 1997; NRC, 1999). Although the impact of pelagic fisheries depletion has been little studied (Botsford *et al.*, 1997; NRC, 1999), it is thought to reduce available food supplies for marine predators including valuable species consumed by humans (Folke and Kautsky, 1989; Fischer *et al.*, 1997). In Europe, recent crashes of North Sea capelin and herring stocks have been attributed to overfishing and may have caused the loss of other wild fish stocks (e.g., cod) and the starvation of seals and seabird chicks (Folke and Kautsky, 1989; Vader *et al.*, 1990). Declining capelin populations in the western Gulf of Alaska are implicated in the decrease of harbor seal and sea lions in the early 1980s (Hansen, 1997). A strong interaction between anchoveta and seabird and mammal populations has also been well documented for the Peruvian upwelling system (Pauly, 1987).

Mollusk culture does not use compound feeds; however, very dense farming of mussels and bivalves in semi-enclosed coastal areas may in exceptional cases reduce fisheries stocks by shortening the linkages comprising the food web. In Rio Arosa, Spain, for example, overgrazing of the phytoplankton population by filter-feeding mussels resulted in zooplankton starvation and the subsequent collapse of the sardine fishery (GESAMP, 1991). This type of ecological disruption, combined with other environmental impacts from culture activities such as conversion of land, habitat modification, and pollution, may lead to net reductions in aquatic productivity over time.

Box 3

Habitat Modification and Biodiversity

Coastal ecosystems, such as mangroves, seagrass beds, and coral reefs, provide habitats and nursery areas for many fish and invertebrate species important to marine fisheries. Almost one-third of the world's marine fish species are found on coral reefs, and fish catch from reefs contributes to human fish consumption approximately 10% globally and much higher in developing countries (Weber, 1993). Harvests of fish, crustaceans, and mollusks from mangrove-lined lagoons and estuaries—excluding mangrove-associated species captured elsewhere—accounted for more than 1% of global fish catch in 1986 (Matthes and Kapetsky, 1988). In the Southeast Asian region, mangrove-dependent species, excluding trash fish, account for approximately one-third of wild fish landings annually (Sasekumar *et al.*, 1994). Moreover, mangroves are closely linked to habitat conditions of coral reefs and seagrass beds through the biophysical interactions in the coastal seascape (Ogden and Gladfelder, 1983; Rönnbäck, 1999).

A positive relationship between fish/shrimp landings and mangrove area has been documented throughout the tropics (Pauly and Ingles, 1986; Rönnbäck, 1999). Mangroves serve as nurseries providing food and shelter for many commercial fish and invertebrate species caught in coastal and offshore fisheries (Robertson and Duke, 1987; Primavera, 1998). In Southeast Asian mangroves, these juveniles belong predominantly to the penaeid shrimps and the fish families Ariidae, Mugilidae, Centropomidae, Serranidae, Carangidae, Leiognathidae, Lutjanidae, and Sciaenidae (Matthes and Kapetsky, 1988; Sasekumar *et al.*, 1994). Other commercial species are associated with mangroves as opportunistic or sporadic foragers. Moreover, mangroves also harbor a wide array of nonmarketed fish, crustacean, and mollusk species, whose subsistence harvest constitutes an important protein source for coastal communities.

Since the 1400s, hundreds of thousands of hectares of mangroves have been transformed into milkfish ponds in Indonesia and the Philippines. In recent decades, shrimp farming has been responsible for a significant share of the conversion of coastal and supratidal areas, e.g., 102,000 ha of mangrove forests in Vietnam in 1983–1987 (Tuan, 1997) and 65,000 ha in Thailand in 1961–1993 (Menasveta, 1997). The loss in wild fisheries stocks due to habitat conversion associated with shrimp farming is potentially very large. If 900 kg of fish, 160 kg of penaeid shrimp, 74 kg of sergestid shrimps, and 25 kg of mangrove mud crabs are harvested annually nearshore for each hectare of mangrove (Rönnbäck, 1999), the 65,000 ha of mangroves converted to shrimp ponds in Thailand result in an annual loss to nearshore Thai fisheries of approximately 58,500 mt of fish, 10,400 mt of penaeid shrimp, 4800 mt of sergestid shrimps, and 1600 mt of mud crabs. With an average productivity of 3000 kg/ha in Thai shrimp ponds (Rosenberry, 1998), 300 g of fish, 53 g of penaeid shrimp, 25 g of sergestid shrimps, and 8 g of mud crabs are lost from capture fisheries for every kilogram of shrimp cultured in these ponds.

These estimates only partially include fish, prawns, crabs, and mollusks that are harvested from mangroves and that contribute to onsite fisheries production. Neither are the off-site fisheries supported by the biophysical interactions between mangroves and other coastal ecosystems such as seagrass beds and coral reefs accounted for in this analysis. Moreover, because penaeid shrimp sales generate most of the revenues from mechanized trawling in developing countries, shrimps (and indirectly mangroves) effectively subsidize commercial harvesting of fish by these vessels (Turner, 1977; Bennett and Reynolds, 1993), including species that do not use mangroves as habitat. The loss to capture fisheries may be higher than fish and shrimp production in mangrove ponds if the full range of ecological effects associated with habitat conversion are accounted for.

Box 4

Wild Capture of Larvae and Spawners for Aquaculture

The farming of shrimp and fish depends on larvae collected from the wild or reared in hatcheries from eggs of wild broodstock or spawners, thereby putting additional pressure on marine fisheries (Fig. 2). The quantities of bycatch associated with such wild catches are directly proportional to the natural abundance of the target species for culture. For example, milkfish *Chanos chanos* constitute only 15% of total fry (excluding great numbers of ambassids, clupeids, and engraulids) in daily inshore collections by seine net (Bagarinao and Taki, 1986). In the Philippines and Indonesia milkfish culture is a major industry which relies heavily on natural seed. The annual utilization of approximately 1.7 billion wild fry for stocking in Philippine milkfish ponds (Bagarinao, 1998) means a corresponding loss of nearly 10 billion fry of other fish species.

In India and Bangladesh, where ponds are predominantly stocked with wild *Penaeus monodon* seed, up to 160 fish and other shrimp fry are discarded for every penaeid shrimp collected from estuarine waters (Primavera, 1998). The same low abundance of larval *P. monodon* applies to adult stages—for trawl catches, 0.14% in the Arafura Sea, Indonesia (Evans and Wahju, 1996), 0.56% in peninsular Malaysia (Ahmad, 1995), and 0.48–0.85% in Guimaras, the Philippines (Primavera, 1995); 0.43–0.68% of commercial fisheries in the Andaman Sea and the Gulf of Thailand (Vibhasiri, 1980); 2.2% of prawn landings in the east coast of India (Rao *et al.*, 1993); and 8.78% of fish corral catches in Batan Bay, the Philippines (Motoh, 1981).

Although the development of hatcheries for cultured shrimp and fish species may have reduced dependence on wild seed (and their mangrove nurseries), it has also increased demand for wild-caught mature (spawners) and immature (broodstock) adults. Because the species is so rare, wild collection of *P. monodon* broodstock and spawners may lead to large amounts of bycatch. Overexploitation of the adults and larvae of both target and incidental shrimp species could be a cause of declining stocks. The staggering quantities of bycatch destroyed by a yearly seed collection of 1 billion *P. monodon* in southeast Bangladesh could have major consequences for biodiversity and capture fisheries production (Deb *et al.*, 1994). In West Bengal, India, where shrimp seed collection constitutes a significant fishery, the contribution of adult shrimp to fisheries landings decreased from 14.4% in 1970–1971 to 8.1% in 1989–1990 (Banerjee and Singh, 1993).

Box 5

Impacts of Wastes, Chemicals, Diseases, and Feral Animals

Nutrient release into marine ecosystems from coastal and nearshore aquaculture operations, especially shrimp and salmon, has been well documented (Gowen and Bradbury, 1987; Briggs and Funge-Smith, 1994; Beveridge, 1996). Similar problems exist in intensive cage culture of carps and tilapia in freshwater ponds, but semi-intensive, extensive, traditional, and polyculture pond systems generally assimilate wastes internally (Beveridge, 1984; Kautsky *et al.*, 1997) and often enhance water quality in the case of integrated carp/agriculture/livestock systems (Kestemont, 1995).

Certain chemicals added to culture systems, such as antibiotics, chlorine, and anti-fouling paints, also harm farmed and wild fish populations. For example, wild fish and shellfish that seek food and refuge near scallop farms in Chile and salmon farms in Scotland contain large amounts of copper and zinc, respectively. Cultured milkfish in the central Philippines have been observed with tin levels of 0.4 mg/kg dry weight (DW) in muscle and 2.3–2.8 mg/kg DW in liver (Coloso and Borlongan, 1999). The acceptable daily intake of total tin for humans is only 0.0005 mg/kg body weight (FAO/WHO, 1991). The high fish levels of tin and triphenyltin (0.3–0.4 mg/kg DW in muscle and 0.6–0.7 mg/kg DW in liver) are striking because application of organotin molluscicides, popular among milkfish pond culturists, had been discontinued (in the wake of a government ban) for 6–12 months in these farms at the time samples were taken.

Further impacts of aquaculture operations on ocean and coastal fisheries can be linked to invasions of feral species (Gausen and Moen, 1991; McKinnell *et al.*, 1997; Gross, 1998; Soto, 1998), spread of diseases and parasites (Rosenthal, 1980; Welcomme, 1988; Chew, 1990), release of nutrients and organic matter (Folke *et al.*, 1994; Costa-Pierce, 1996), and discharge of chemicals and antibiotics (Beveridge, 1984; Alderman *et al.*, 1994; Weston, 1996; GESAMP, 1997). These factors—combined with increasing reliance on external feeds, habitat conversion, and bycatch—negatively affect fisheries through qualitative and quantitative changes in spawning and nursery grounds, species composition, and food webs of adjacent ecosystems (Fig. 2).

In some cases, fish culture reduces wild fish stocks through the introduction and invasion of exotic species. Tilapias have had a long history of purposeful (for aquaculture and fisheries) and accidental introductions to 90 countries and territories. Limited information available on the environmental impacts of these introductions indicates that the presence of tilapias is only one of many factors contributing to the decline of native fish species in some waters in the Philippines and Bangladesh (Pullin *et al.*, 1997). Atlantic salmon from culture facilities have escaped within the geographic range of wild Atlantic salmon, as well as in Pacific waters, and are now found as far north as the Bering Sea and as far south as Chile. Increasing evidence suggests that these escapes may have direct genetic impacts on wild populations through hybridization (Gross, 1998). Larger numbers of escapes also increase the likelihood of hybridization between farmed Atlantic salmon males and wild females in populations that are locally endangered or close to extinction (Slaney *et al.*, 1996; Gross, 1998). In addition to consequences for the population gene pool and fitness, there are many potential ecological impacts associated with feral fish. Atlantic salmon may compete extensively with wild salmon species for food and space, disturb native spawning sites, and introduce new diseases and parasites into wild populations (Beveridge *et al.*, 1994). Numerous wild salmon stocks in Norwegian rivers have been infested by an ectoparasitic trematode through the release of juvenile farmed salmon (Folke and Kautsky, 1989).

Diseases are prevalent in other aquaculture species as well, especially marine shrimp. Worldwide transfers and introductions of the few preferred culture species, including *P. monodon*, *Litopenaeus vannamei*, and *Marsupenaeus japonicus*, were numerous in the early decades of commercialized shrimp culture. At the peak of Taiwanese shrimp production in 1982–1986, yearly imports from Southeast Asia of more than 100,000 live *P. monodon* broodstock supported hatchery production (Chin, 1988). Such introductions and transfers may lead to competition with endemic fauna, genetic introgression with local fauna, and introduction of pathogens and parasites (Beardmore *et al.*, 1997).

Recently, the introduction of shrimp postlarvae and broodstock from areas affected by the whitespot syndrome virus (WSSV) and Taura syndrome virus (TSV) was often followed by the rapid spread of these major pathogens throughout most of the shrimp-growing regions in Asia and Latin America, respectively. A native of Asia, where it has caused multimillion-dollar shrimp crop losses, the WSSV has been detected in wild and cultured shrimp (*L. setiferus*, *L. vannamei*, *L. stylirostris*, and *Farfantepenaeus duorarum*) in Texas and South Carolina (Lightner *et al.*, 2000). The virus was probably introduced by release of untreated wastes from plants processing imported Asian shrimp into coastal waters and by use of imported shrimp as bait in sports fishing or as fresh food for rearing other aquatic species (Lightner *et al.*, 1997). Another major shrimp virus, the infectious hypodermal and hematopoietic necrosis virus (IHHNV), is believed to have been introduced to the Americas from Asia through the importation of live *P. monodon* in the early 1970s (Lightner *et al.*, 1999). In the Philippines, IHHNV prevalence in various wild populations of the giant tiger prawn has been correlated with shrimp culture intensification and mangrove status (Belak *et al.*, 1999). Lower viral incidence in wild shrimp has been found in sites with primary mangroves and no major aquaculture industry, whereas higher levels have been observed in areas with intensive shrimp farms and severely degraded mangroves. Wild populations had a significantly lower overall IHHNV incidence of 51% compared to total infection in captive *P. monodon* reared from second- and third-generation hatchery fry.

In contrast to shrimp and salmon, comparatively few diseases have been reported for carps, tilapia, and milkfish, particularly from extensive and other low-density culture systems. The current trend toward intensification in rearing ponds and cages, however, may create stressful conditions through deterioration of water quality, excessive stocking, and polluted water inflow that predispose the fish to disease. For example, pond discharges of nutrients from feces and uneaten food may alter nutrient and organic matter concentrations, stimulate blooms of toxic algae, and generate self-pollution that is conducive to disease outbreaks.

Box 6

The Ecological Footprint: A Tool for Assessing Resource Use in Aquaculture

To reduce the risk of resource constraints and biodiversity impacts, we need to shift to aquaculture production systems that consume less resources and release levels of wastes within the assimilative capacity of the environment. We also need to recognize and manage nature's life support on which economic development and human welfare depends. One way to identify human demands for natural resource and ecosystem services is by estimating the functional ecosystem area—the ecological footprint—required to support human activities.

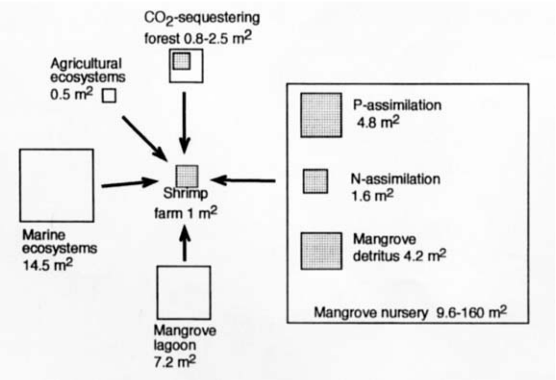


FIGURE 3 Ecosystem support areas required to sustain a semi-intensive shrimp farm in a coastal mangrove area of Colombia (square meters of support area needed per square meter of pond area).

The concept has proven useful in illuminating the nonvalued (or monetized) and often unrecognized work of nature that forms the basis for economic activities such as industrial aquaculture. When problems beset aquaculture operations, solutions focus on the pond or cage unit, and it is not considered that the farm is part of a much larger ecosystem with which it interacts. Surrounding ecosystems provide the feed, seed, clean water, and other required resources and services including waste assimilation. This unvalued work of nature sets the limits to culture levels without compromising biodiversity or causing pollution or disease problems. Many aquaculture developments, such as intensive shrimp farming, have encountered problems or even failed because they exceeded the carrying capacity of the environment.

A detailed illustration of the footprint concept is provided by a study on semi-intensive shrimp farming in a coastal mangrove area in Colombia. This study estimated that the spatial ecosystem support or footprint required to produce food inputs, nursery areas, and clean water as well as to process wastes was 35–190 times the surface area of the farm (Fig. 3) (Larsson *et al.*, 1994). The mangrove nursery area required to produce the shrimp seed for stocking was the largest support system covering up to 160 times the pond area. If located close to the farm, the same mangrove area could also supply natural food inputs (4.2 m^2 per m^2 shrimp pond area) and absorb polluting nutrients (2–22 m^2 per m^2 pond area) in the farm effluents. Feed pellets form a major input to a shrimp farm, and a marine area of 14.5 m^2 was needed to catch the fish, and an additional agricultural area of 0.5 m^2 for the vegetable ingredients used in feed pellet manufacturing was also needed. Finally, 7.2 m^2 was needed for providing clean lagoon water to the ponds, and 0.8–2.5 m^2 of forest area per square meter shrimp pond area was needed to sequester the CO_2 of fossil fuel burning at the farm.

Footprint size will change with farming intensity, i.e., a higher stocking density will require more food inputs and produce more wastes (Folke and Kautsky, 1992). Pressure on local ecosystems can be reduced to some extent by importing some inputs (e.g., feeds) from other areas and by investing in shrimp hatcheries, the costs of which can be absorbed in the budget. Although producing seed in hatcheries reduces large amounts of shrimp and fish larvae bycatch that would otherwise be recruited to fisheries (Primavera, 1998; Rönnbäck, 1999), it also increases demand for wild-caught spawners and broodstock.

Other services, however, such as clean water supply and waste assimilation, must be located close to the farming area. This may be no problem up to a certain level of farming intensity, but the whole operation may collapse when the dynamic carrying capacity of the local environment is exceeded unless extensive and costly pipelines and water treatment facilities are built. The footprint concept provides an early warning device when the level of carrying capacity is being approached. Integrated farming technologies that recirculate resources and wastes within the farm may be one way of reducing the footprint (see Box 7).

Because the surrounding mangrove resources and the lagoon water supply were becoming limiting, the Colombia study recommended no further shrimp farm expansion in the area and preservation of remaining mangroves. According to available information, shrimp farming has not expanded much since the study was made in 1992 and the farms have not been struck by disease problems as severely as has occurred in many other areas.

TABLE II

The Ecological Footprint of Seafood Production

	Coastal and marine support areas	
Activity	Resource production	Waste assimilation
Salmon cage farming, Sweden	40,000–50,000	—
Tilapia cage farming, Zimbabwe	10,000	115–275
Salmon tank system, Chile	—	16–180
Shrimp farming (semi-intensive) Colombia	34–187*	—
Shrimp farming (semi-intensive) Asia	—	2–22
Mussel rearing, Sweden	20	—
Tilapia pond farming, Zimbabwe	0	0

Values are area of footprint per area of activity (m2/m2) (Data from Folke *et al.* 1998).

* Calculated from level of fisheries. If recalculations are made to the primary production base as for other activities the support area will increase to about 10,000.

Ecological footprints have been established for many aquaculture commodities (Table II). The data illustrate that the coastal and marine support areas needed for resource inputs and waste assimilation range from as large as 50,000 times the cultivation area for intensive salmon cage farming to negligible for extensive and semi-intensive tilapia pond farming. The latter uses discards from fisheries and agricultural waste as the sole inputs (Table II).

In its current stage, the concept is useful for communicating the importance of viable ecosystems to farmers and policymakers. Although it is a static measure in its current form, the footprint concept reveals the hidden requirements for ecosystem support and locates (or situates) fisheries and aquaculture within a larger ecosystem framework. Human activities that seem separate from nature would not function without ecosystem support.

Box 7

Integrated Aquaculture

Commercial aquaculture is usually practiced using large monocultures. Like all types of monocultures, large-scale, and especially intensive aquaculture will result in negative environmental impacts (see Box 5) and make the surrounding waters less suitable or even unusable for other purposes, including the further culturing and harvesting of natural stocks. The problem with these types of cultivations is their throughput characteristics (Daly and Cob, 1989; Folke and Kautsky, 1992), in which large amounts of wastes are released beyond the capacity of the environment to recycle. The fact that treatment of effluents usually involves a high degree of technology and therefore high costs implies that release of untreated water is the rule rather than the exception, especially in many developing countries. Traditional pond cultures of herbivorous fish species (e.g., carps in China), however, have been viable for centuries and their mere existence is proof of sustainable integrated farming systems. Here, raising poultry and livestock is integrated with fish culture, and the principal linkages between raising poultry or livestock and aquaculture are animal manure and other agriculture waste products. Compost is used to fertilize the pond water for proliferation of natural organisms as natural feeds for fish from juvenile to adult. Such systems utilize fish species low in the food web, making efficient use of inputs, and generate little waste. The production also adds, at relatively minor environmental and social expense, to net fish supplies at local and regional levels.

However, increased world demand for seafood products cannot be met by such traditional extensive production systems, and as previously mentioned current modern, intensive aquaculture practices cannot be regarded as a suitable alternative. As an alternative to monocultures and high-tech pollution treatment solutions, the practice of ecological engineering (Mitsch, 1989) may offer solutions to increasing production without deteriorating the environment. The concept has recently gained new interest in an aquaculture context with negative environmental effects being remedied by species integration for nutrient trapping or recirculation. Wastewater from land-based fish and shrimp cultivation and open fish cage cultivation has proven to be a suitable nutrient source for culture of seaweeds and bivalves. The systems have different characteristics depending on how waste is emitted to the surrounding ecosystems (Troell *et al.*, 1999), implying that one solution applied in one type of system may not be optimal in another.

Filter feeders such as mussels and oysters have successfully been integrated with fish aquaculture (Jones and Iwama, 1991; Shpigel *et al.*, 1993) and the technique is also being practiced in shrimp aquaculture (Hopkins *et al.*, 1993). Filter feeders in such an integrated system benefit from release of particles, from waste feed and feces from the farm, and from the stimulation of bacterial production and phytoplankton cells. In addition to generating additional income for the farmer, such an integrated system will also secure income by resulting in a more diversified production (Troell *et al.*, 1999). In addition to output of particulate wastes, aquaculture also releases dissolved nutrients, and generally less than one-third of the nutrients added through feed will be removed through harvest in intensive fish and shrimp farming (Gowen *et al.*, 1991; Primavera, 1994; Briggs and Funge-Smith, 1994). Many studies have verified that wastewater from intensive and semi-intensive tank or pond mariculture systems is suitable as a nutrient source in seaweed production (Neori *et al.*, 1991; Buschmann *et al.*, 1994, 1996; Jiménez *et al.*, 1994; Krom *et al.*, 1995; Neori, 1996; Troell *et al.*, 1999), and that integration with seaweeds can significantly reduce the loading of dissolved nutrients to the environment (Lin *et al.*, 1993; Phang *et al.*, 1996). The choice of commercially attractive seaweed species also increases profitability for the farmer in integrated cultivation. In more open culture systems such as cage farming, the continuous exchange of water makes waste disposal difficult to control, but studies have shown beneficial effects from the integration of seaweeds with such culture (Hirata *et al.*, 1994; Petrell *et al.*, 1993; Troell *et al.*, 1997).

The development of integrated techniques in which seaweeds and filter feeders are used as biofilters is in its infancy and further research on ecologically sound production systems is needed. Even though the development of closed, intensive recirculating systems seems to be of high priority (e.g., shrimp culture), the use of biofilters could add multiple benefits to such systems. Integrated farming may thus enhance biodiversity, both by the practicing of polyculture and by lowering pollution pressure on the environment.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • ENERGY FLOW AND ECOSYSTEMS • FISH STOCKS

Bibliography

Baird, D. J., Beveridge, M. C. M., Kelly, L. A., and Muir, J. F. (Eds.) (1996). *Aquaculture and Water Resources Management*. Blackwell, Oxford.

Beardmore, J. A., Mair, G. C., and Lewis, R. I. (1997). Biodiversity in aquatic systems in relation to aquaculture. *Aquacult. Res.* **28**(10), 829–839.

Beveridge, M. C. M. (1996). *Cage Aquaculture* (2nd ed.). Fishing News Books, Oxford.

Beveridge, M. C. M., Ross, L. G., and Kelly, L. A. (1994). Aquaculture and biodiversity. *Ambio* **23**, 497–502.

Beveridge, M. C. M., Ross, L. G., and Stewart, J. A. (1997a). The development of mariculture and its implications for biodiversity. In *Marine Biodiversity: Patterns and Processes* (R. F. G. Ormond, J. Gage, and M. Angel, Eds.), pp. 372–393. Cambridge Univ. Press, Cambridge, UK.

Beveridge, M. C. M., Phillips, M. J., and Macintosh, D. C. (1997b). Aquaculture and the environment: The supply and demand for environmental goods and services by Asian aquaculture and the implications for sustainability. *Aquacult. Res.* **28**, 101–111.

De Silva, S. (Ed.) (1998). *Tropical Mariculture*. Academic Press, London.

Folke, C., and Kautsky, N. (1989). The role of ecosystems in the sustainable development of aquaculture. *Ambio* **18**, 234–243.

Folke, C., and Kautsky, N. (1992). Aquaculture with its environment: Prospects for sustainability. *Ocean Coastal Management* **17**, 5–24.

Folke, C., Kautsky, N., Berg, H., Jansson, Å., and Troell, M. (1998). The ecological footprint concept for sustainable seafood production—a review. *Ecol. Appl.* **8**, 63–71.

Lightner, D. V., Redman, R. M., Poulos, B. T., Tang, K. F. J., Nunan, L. M., Mari, J. L., Pantoja, C. R., Wang, Q., Zhou, J. L., Durand, S., Hasson, K. W., and White, B. L. (2000). The penaeid shrimp viruses TSV, IHHNV, WSSV and YHV: Current status in the Ameridas, available diagnostic methods and management strategies. *J. Appl. Aquacult.*, in press.

Matthes, H., and Kapetsky, J. M. (1988). Worldwide compendium of mangrove-associated aquatic species of economic importance. *FAO Fish. Circ.* **814**, 1–236.

National Research Council (1999). *Sustaining Marine Fisheries*. National Academy Press, Washington, D.C.

Naylor, R., Goldberg, R., Mooney, H., Beveridge, M., Clay, J., Folke, C., Kautsky, N., Lubchenco, J., Primavera, J., and Williams, M. (1998). Nature's subsidies to shrimp and salmon farming. *Science* **282**, 883–884.

Naylor, R., Goldberg, R., Primavera, J., Kautsky, N., Beveridge, M., Clay, J., Folke, C., Lubchenco, J., and Troell, M. (2000). Effects of aquaculture on world fish supplies. *Nature* **405**, 1017–1024.

Primavera, J. H. (1998). Tropical shrimp farming and its sustainability. In *Tropical Mariculture* (S. deSilva, Ed.), pp. 257–289. Academic Press, London.

Pullin, R. S. V., Rosenthal, H., and Maclean, J. L. (Eds.) (1993). Environment and aquaculture in developing countries. ICLARM Conference Proceedings No. 31. ICLARM, Manila.

Troell, M., Halling, C., Nilsson, A., Buschmann, A., Kautsky, N., and Kautsky, L. (1997). Integrated open sea cultivation of the seaweed *Gracilaria chilensis* (Gracialariales, Rhodophyta) and salmonids for reduced environmental impact and increased economic output. *Aquaculture* **156**, 45–61.

Troell, M., Rönnbäck, P., Kautsky, N., Halling, C., and Buschmann, A. (1999). Ecological engineering in aquaculture. The use of seaweeds for removing nutrients from intensive mariculture. *J. Appl. Phycol.* **11**, 89–97.

Weston, D. P. (1991). The effects of aquaculture on indigenous biota. *Adv. World Aquacult.* **3**, 534–567.

ARACHNIDS

Jonathan A. Coddington* and Robert K. Colwell†
**Smithsonian Institution; †University of Connecticut*

GLOSSARY

booklungs One to four pairs of abdominal respiratory organs consisting of a thin, multifolded membrane (the book's "pages") over which blood circulates and that is open to an air-filled cavity on the outside, itself open to the exterior via a spiracle. Gases passively diffuse back and forth across the membrane.

chelicerae (chelate) The first pair of preoral appendages. They are at most three segmented, usually two, and usually the distal segment acts against the penultimate to grab or hold prey or objects. If the basal segment has a finger-shaped outgrowth against which the distal segment operates, the chelicerae are chelate (as in scorpions or harvestmen). If not, the chelicerae are subchelate, as in spiders and tailless whip scorpions. In parasitic mites the chelicerae are modified into piercing stylets.

monophyly A true, historical, evolutionary lineage consisting of an ancestor and all of its descendants; defined by shared, derived characters.

ovoviviparous Young are born alive, but the mother simply retains eggs within her body until they hatch.

paraphyly A group consisting of an ancestor and only some of its descendants. Defined by primitive characters.

pedipalps The second pair of preoral appendages. They are multisegmented and primitively leg-like. They may be raptorial or sensory (like antennae) or used as walking legs.

phoresy A method of long-range dispersal in which the dispersing animal attaches itself to another animal (e.g., beetle, wasp, or bird) that carries the disperser along with it until the disperser drops off or disembarks.

polyphyly A group in which the most recent common ancestor of the included taxa is excluded from the group. Defined by convergent, nonhomologous characters.

spermatophore A chitinous container produced by the male to hold sperm. It may be attached to the substrate for the female to find or passed to the female from the male during mating.

spinnerets Usually three, rarely four pairs of modified terminal abdominal appendages in spiders bearing one to hundreds of hollow spigots from which silk is drawn.

tracheae A system of hollow, branched or unbranched air-conducting tubes used for respiration, opening via abdominal spiracles. They may or may not extend into the cephalothorax or legs.

trichobothria Long, delicate, slender setae set in broad, shallow innervated sockets in the cuticle. Trichobothria are sensitive to vibration or near-field air movement and are a major sense organ of arachnids.

TABLE I

Latin name	Common name	No. of Families	No. of Genera	No. of Described species
Arachnida	Arachnids	648	9241	±92,680
Araneae	Spiders	108	3200	±37,000
Palpigradi	Micro-whip scorpions	2	6(5)	80
Uropygi	Whip scorpions	1	16	101
Amblypygi	Whip spiders, tailless whip scorpions	5	20	126
Schizomida	Schizomids	2	31	195
Solifugae	Wind spiders	12	153	1,065
Pseudoscorpiones	Pseudoscorpions	24	430	3,100
Ricinulei	Ricinuleids	1	3	53
Opiliones	Harvestman	44	1554	±4,500
Scorpiones	Scorpions	18	156	1,260
Acari	Mites, ticks	431	3672	45,200

[a] Modified from Adis and Harvey (2000), with data for the Acari from Walter and Proctor (1999).

Arachnida is a class of the huge phylum Arthropoda. Familiar arachnids are spiders, scorpions, ticks, mites, and harvestmen, but arachnids include many lesser-known terrestrial arthropod groups as well.

I. OVERVIEW OF ARACHNIDA

The known diversity of arachnids is approximately 640 families, 9000 genera, and 93,000 species (Table I), but there are many thousands of new mite and spider species still undescribed and hundreds to thousands of undescribed species in the remaining orders. Together with the marine horseshoe crabs (Xiphosura) and sea spiders (Pycnogonida), arachnids comprise the arthropod subphylum Chelicerata, named for the characteristic first pair of preoral appendages, the chelate, or pinching, mouthparts. In some arachnid groups the chelicerae are further modified into venomous fangs or piercing stylets to suck body or plant fluids. Arachnids are the only terrestrial chelicerates. Along with the insects, arachnids are by far the most species-rich, abundant, and widely distributed terrestrial arthropods. Acarologists (scientists who study mites and ticks) estimate that there may be as many mite species as beetles, implying that total extant arachnid diversity may exceed 1 million species. Arachnids are an important component of every terrestrial ecosystem, but, apart from several specialized mite lineages, none are aquatic or marine. Although most arachnids share many ancestral similarities in body plan and lifestyle, many extremely specialized groups exist, especially among the mites (Acari).

Although arachnids are commonly mistaken for some sort of peculiar insect, the groups are quite distinct and only distantly related. Arachnids have four pairs of walking legs rather than three (except the young stages of mites and the related Ricinulei), only two (not three) major body parts, and simpler chelate mouthparts rather than the more complex feeding apparatus of insects. The anterior body part is specialized for locomotion and the posterior for digestion and reproduction. Arachnids lack the wings, antennae, and compound eyes usual in insects. In many groups the first pair of walking legs (e.g., amblypygids, uropygids, schizomids, solifuges, palpigrades, and many mites) are elongate and function in much the same way as insect antennae. Figure 1 depicts a consensus view of the position of arachnids in the arthropod evolutionary tree.

Arachnids and hexapods (which insects dominate) differ in fundamental ways, possibly because their marine ancestors were already distinct lineages in the Silu-

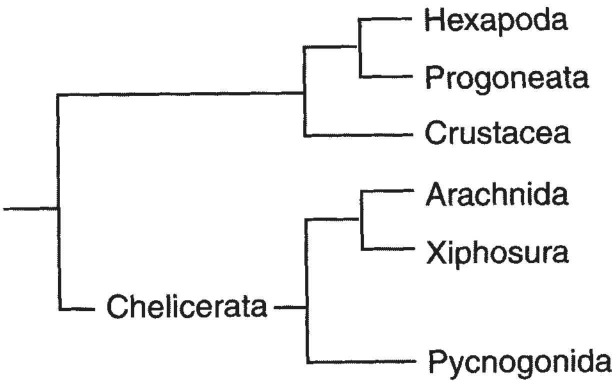


FIGURE 1 Phylogenetic position of Arachnida among arthropods.

rian and both colonized land independently. They therefore solved the fundamental challenges of terrestrial existence (support, breathing, water balance, reproduction in dry environments, and nitrogenous waste management) in different ways. The arachnid skeleton is hydraulic; arachnids, except scorpions and pseudoscorpions, lack extensor muscles at key joints. Instead, the animal pumps blood into the limb to extend it. The basic arachnid uses two (or four) pairs of book lungs (gas-permeable, gill-like membranes with blood on one side and open to the air on the other) to exchange carbon dioxide for oxygen rather than tracheae. Many arachnids possess rudimentary tracheal systems that supplement or replace book lungs, but ventilation is passive, not active. Unlike insect tracheae, arachnid tracheae generally do not ramify throughout the entire body or penetrate inside body cells. For all these reasons, insect tracheae more efficiently deliver oxygen directly to tissue, and insects in general can lead more energy-intensive lifestyles than arachnids. Arachnids have significantly lower metabolic rates than other terrestrial arthropods, especially insects. Whereas male insects transfer sperm directly to the female using an intromittent organ, in most arachnids males either ejaculate onto a special structure and carry the sperm mass with a specialized appendage until the female is encountered or they deposit the sperm mass in a specially built receptacle (spermatophore) fixed to the substrate, which the female picks up. Major exceptions include the harvestmen, astigmatic mites, and spider mites, which transfer sperm to the female by means of an intromittent organ, and the water mites (Prostigmata: Hydracarina), which transfer sperm directly by opposing the male gonopore to the female gonopore.

Arachnids are peculiar among animals in using guanine (three nitrogen atoms per molecule) as well as the much more common uric acid (two nitrogen atoms per molecule) to eliminate nitrogenous wastes. Insects have compound eyes, which provide relatively excellent vision. Arachnids lost compound eyes early in their evolutionary history but retain usually one to five (commonly four) pairs of simple eyes, much inferior in acuity to compound eyes. Schizomids, palpigrades, ricinuleids, mites, and other mainly litter-dwelling arachnid groups are nearly always blind. Vision is much less important to arachnids than vibration. Many structures (slit sense organs, trichobothria, and lyriform organs) are specialized to detect minute vibrations and slight air currents.

Arachnids are also peculiar in that species of most arachnid groups digest food externally. They have a strong pumping stomach that rhythmically vomits and sucks digestive juice through a preoral cavity formed by the basal articles of the pedipalps back and forth over their prey. The process continues until only the hard, indigestible parts of the prey remain. Only liquids or very small particles are actually ingested. Some major groups of arachnids have internal digestion, however. Opilionids and several groups of mites are particulate feeders, and parasitism of plants, vertebrates, and invertebrates has arisen repeatedly in the mites. External digestion is a major obstacle to life in fresh water or the sea. Arachnids are also notable for their ability to withstand starvation. Fasts of weeks or even months are routine for larger arachnids. Some scorpions and mygalomorph spiders live for more than 1 year without food, and adult soft ticks (Argasidae) can survive for years without feeding.

A. Reproduction and Growth

Like all arthropods, arachnids grow by molting their exoskeletons and expanding the larger skin beneath with blood pressure before it hardens into the usual tough covering. The number of molts to maturity varies widely between 3 and 10–12; five is perhaps the most common. Life spans also vary greatly. The majority of mites and spiders live less than 1 year, but several years is common among the larger forms, and mygalomorph spiders can live 20–30 years in captivity. Some arachnids cease molting at adulthood but others continue to molt periodically until death.

The ancestral reproductive pattern of sperm transfer is via spermatophore, modified in spiders, harvestmen, mites, and ricinuleids. Except among permanently social species and many mites, the sex ratio is equal, and parthenogenesis is rare. In the vast majority of species, males and females meet only to mate; cohabitation and parental care are uncommon. Nevertheless, various spiders, scorpions, schizomids, uropygids, and amblypygids may carry and feed their young, and in some harvestmen the male cares for the eggs in a specially built nest. Pseudoscorpion females nourish eggs with secretions from their bodies, and scorpions bear only live young. Females commonly guard their eggs until they hatch, but the young are usually abandoned soon thereafter. Mites are more diverse in reproductive strategies than nonacarine arachnids.

B. Ecology

Despite huge numbers of species (Table I), arachnid biology is coherent in many ways (even though exceptions to nearly every generalization exist). Most arachnid orders consist of fluid-feeding predators, and predation still dominates these groups today. However, opilionids and two of the three orders of mites are particulate-feeders on detritus, fungi, and small invertebrates. Additionally, parasitism of vertebrates, invertebrates, and plants has arisen numerous times within the Acari and radiations within these lineages account for most of the 45,000 described species of mites.

Non-acarine arachnids tend to be at the top of the terrestrial invertebrate food chain wherever they occur. At one site in Israel, mites comprised 35% of the total soil arthropod population; in the Amazon ranges from 35 to 55% have been reported. Some harvestmen eat dead or decaying animal or plant material. Arachnids are generally nocturnal, despite numbers of diurnal harvestmen, spiders, and mites. Nocturnal forms hide in dark crevices and burrows during the day; several orders are morphologically specialized to inhabit small spaces. With the exception of the wind spiders (Solifugae), arachnids tend to be torpid and sedentary—none fly, for example, nor do any move constantly like ants or other active insects. The basic arachnid forages with a "sit and wait," solitary strategy. They move rarely and wait for prey to encounter them. Prey is then seized with a quick strike and immobilized. Highly organized social systems are known only in a few spiders, but loose aggregations (spiders, harvestmen, and pseudoscorpions) are not uncommon, usually in response to high prey density or limited refugia. Most arachnids are well adapted to last for long periods (weeks or months) without food; in the laboratory some have survived for more than 1 year. Only jumping spiders among arachnids have notably good vision; otherwise, they orient primarily via vibrations and touch.

Arachnids occupy all terrestrial habitats—deserts, forests, tundra, grasslands, mountaintops, soils, litter, caves, etc. Hydrachtinid mites (approximately 5000 species) are important components of most freshwater ecosystems; other mites are parasites of marine organisms, whereas others inhabit marine sediments, including the deepest oceanic trenches. Otherwise, arachnids are exclusively terrestrial. A few groups, such as scorpions and wind spiders, conserve body water as well as any arthropod and thus tend to dominate in deserts. The majority needs moist conditions to survive. Schizomids, palpigrades, and ricinuleids are apparently restricted to the interstices of moist tropical and subtropical leaf litter or equally constant and moist habitats.

C. Phylogeny and Taxonomy

The most commonly encountered arachnids are spiders, scorpions, harvestmen, and mites, but the class contains seven smaller groups of terrestrial arthropods less familiar to the general public (Fig. 1). The largest and heaviest arachnid is the African scorpion *Pandinus imperator*, which may reach a length of 18–20 cm. The smallest are perhaps the gall mites at 80 μm. No group of arachnids is well-known by vertebrate, butterfly, or vascular plant standards. Popular manuals are available for only a smattering of the most common species of spiders in Europe, North America, and Japan; all others require technical literature and specialist knowledge to identify. Myriad species are undescribed and undiscovered; at best, the approximately 93,000 known arachnid species are but one-third of the probable total, and probably much less. Most undescribed arachnid species are mites.

The study of arachnids is called arachnology. The principal international scientific society for nonacarines has approximately 600 members but many more belong only to regional societies. The taxonomy of arachnids is still a monumental task and an obstacle to better ecological and biotechnological understanding of arachnids, but the number of arachnid taxonomists is small and decreasing, and new students are not being trained. There are no comprehensive arachnology texts appropriate for university teaching (three exist for Acari), although modern "biologies" are available for spiders, scorpions, solifuges, pseudoscorpions, and various aspects of mite biology.

D. Paleontology

Arachnida were among the earliest terrestrial animals. The marked similarity between fossil and recent forms in overall body plan and morphology suggests few changes over hundreds of millions of years. The earliest sites for terrestrial arachnids are Early Devonian (400 Ma) and Late Silurian (414 Ma); the extinct arachnid order Trigonotarbida figures prominently, but mites are also present. The fossil record of all arachnids is comparatively poorly known. The 13 living and extinct orders are still known from less than 50 major time horizons since the Silurian; gaps remain more common than fossils. Arachnids seem to have invaded land in the Silurian and reached a pinnacle of ordinal diversity by the Carboniferous. The latest dates for the extinct orders Trigonotarbida and Phalangiotarbi are Permian and Carboniferous, respectively. No order seems to have succumbed to the end Cretaceous event that eliminated the dinosaurs. The earliest arachnid fossils are aquatic scorpions from the Late Silurian. Spiders, pseudoscorpions, terrestrial scorpions, and mites are known from the Devonian. The scorpion-like Eurypterida, which are sister to true Arachnida, also may have become extinct at the Permo–Triassic boundary; the youngest fossils are also Carboniferous. The phylogeny in Fig. 2 implies that many arachnid clades must predate all scorpion fossils; an alternate opinion, based on less

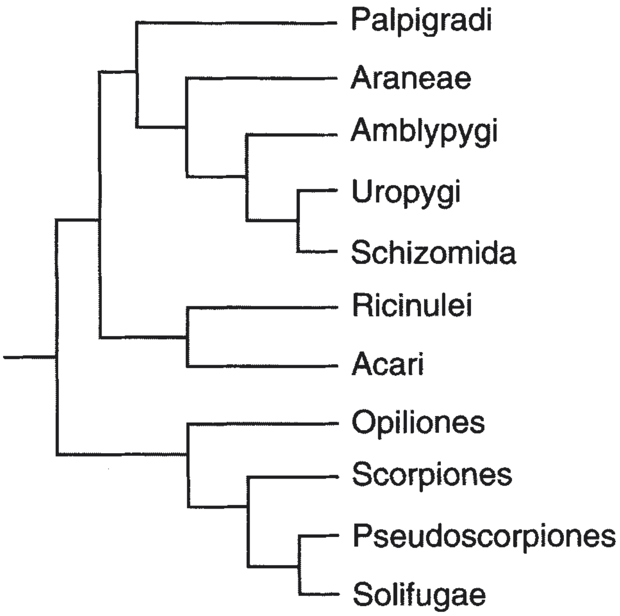


FIGURE 2 Phylogeny of arachnid orders.

evidence, is that scorpions instead are sister to eurypterids, and that the phylogeny of Fig. 2 will be shown to be incorrect. Small size may explain the rarity of schizomid and palpigrade fossils, but mites in the Devonian show that small arachnid fossils can persist. Given the sparse record, it may well be that all arachnid orders were distinct and essentially modern in appearance by end Silurian–Early Devonian times. The most detailed arachnid fossils are "inclusions" in amber, which is fossilized gum or tree sap. The organisms became caught in the wet sap and were fossilized along with the sap. Cenozoic amber arachnid fossils 40–80 million years old can mostly be placed in modern genera; in some cases, the fossils are difficult or impossible to distinguish from modern species. Ecological relations are also preserved: Parasitic mites occur on their insect hosts, for example, thus documenting the antiquity and diversity of mite parasitism.

II. ARANEAE

Araneae is the second largest arachnid order with 108 families, 3200 genera, and approximately 37,000 species described (second to mites). Spiders are distinguished from other arachnids by their silk-producing spinnerets at the end of the abdomen and the prosomal poison glands exiting through their chelicerae modified as fangs (Fig. 3). Spiders are among the very few animals that use silk throughout their entire lives. A narrow stalk joins the abdomen to the prosoma, allowing great flexibility and precise orientation of the abdominal spinnerets. Pedipalps are leg-like and short. Colors are predominantly dull tans, browns, and blacks, but spiders are occasionally very colorful, even iridescent. The abdomen of most spiders shows no trace of ancestral segmentation, unlike that of other arachnids. The Neotropical tarantula *Theraphosa leblondi* (Theraphosidae) is the largest spider at about 10 cm in body length. The smallest spiders are the tiny orb-weaving Symphytognathidae; adults are less than 1 mm long.

FIGURE 3 Spider (Araneae: Ctenidae, *Cupiennius* sp.) (photograph by Jonathan A. Coddington).

A. Ecology

Spiders are very abundant. One calculation estimated 5 million animals per hectare in an English meadow. Another found 29,000 per cubic meter in an English sewage treatment plant. These are extreme values, but "average" nondesert habitat probably supports at least 1 and as many as 800 spiders per square meter. Many spider species disperse by ballooning. Ballooning spiders spin a silk line until it is caught by the wind and lifted aloft—potentially for hundreds or thousands of kilometers. One study estimated that 216,000 spiders per hectare may balloon into tilled fields during the growing season. Point diversities per hectare may vary from 100 species in moderate temperate zones to 600 or more in the wet tropics. Boreal diversities are less, perhaps 20–50 species per hectare.

Spiders are the only animals to use silk throughout their lives. Silk is one of the strongest and toughest natural fibers known and compares favorably with the best man-made filaments. Eight different kinds of spider silk have been discovered, but the maximum made by any single species is seven. Orb web spinners, for example, make two varieties of stiff, tough silk for weight-bearing structural fibers and safety lines, cement to fix silk to itself and substrate, sticky silk or glue to capture prey, rubbery silk to carry the sticky silk in the web, specialized silk for eggsacs, and thin, weak silk spun as multiple fibers to wrap prey, cradle eggs, and for other general purposes. All spiders are capable of spinning silk as soon as they leave the eggsac, and all make at least safety lines ("draglines") and the cement to attach them to substrate. Approximately half of spider species spin webs to capture prey. Web architectures are taxonomically specific and provide many clues for reconstructing spider evolution. The remaining spiders are ambush predators such as the crab spiders, which lurk inside flowers to attack pollinators, tube or retreat dwellers that forage in the very limited area at the burrow mouth, or vagabond predators such as wolf, ground, or jumping spiders.

B. Reproduction and Growth

Reproduction in spiders requires the male to ejaculate sperm onto a specially constructed sperm web. He then sucks up the sperm into specialized sperm transfer organs at the pedipalpal tips. Only adult males have such structures; their form is usually species-specific. With his palps filled, the male then searches for females. Females are usually more sedentary than males. Life is short for adult males, both because of predation and because they eat little as adults. The duration of courtship and copulation varies from seconds to days, but it ends with the insertion of the male pedipalp and transfer of sperm into the female gonopore. Females store the sperm and sometime later construct a silken eggsac into which 1–2500 eggs are placed. Eggs are fertilized only as they exit the female's body. Parental care is rare and highly variable, ranging from simple guarding of eggsac to actively feeding the babies. In at least one crab spider the female dies as the juveniles emerge from the eggsac, which then eat their mother's body. About 20 species of spiders are extremely social. Males are rare (1:40–100), generations overlap, food is shared, and prey capture and brood care are cooperative. Colony sizes range from a dozen to several thousand individuals.

C. Phylogeny and Taxonomy

Spiders are the seventh largest zoological order on Earth (after Coleoptera, Hymenoptera, Lepidoptera, Diptera, Hemiptera, and Acari), and of these they are the only one for which all taxonomic literature is fully cataloged. Catalogs greatly facilitate all kinds of research because scientists can easily determine the current taxonomic status and history of any described species and thus decide whether a given specimen belongs to a described species or not. At higher taxonomic levels (approximately, family) a basic, first-draft phylogeny is nearly complete. At the species level the easiest species-specific characters are found in the male and female genitalia. Species boundaries in spiders are generally clear-cut. However, the species taxonomy of spiders is based overwhelmingly on morphology. If more costly and sophisticated molecular methods were routinely applied, the number of distinguishable spider "taxa" would certainly increase.

The fundamental phylogenetic division in spiders is between the primitive mesotheles (spinnerets towards the middle of the abdomen) and the opisthotheles (spinnerets terminal). Within opisthotheles there are again two basic groups, the mygalomorphs (tarantulas and their allies) and araneomorphs (so-called "true" spiders). Mesotheles and mygalomorphs are not particularly diverse at the species level; araneomorphs currently include 94% of all known spider species and this disparity will certainly increase. Within araneomorphs the basal taxa are a few relictually distributed families in north and south temperate regions. Araneomorph haplogynes are diverse, but again comprise relatively few species. The araneomorph Entelegynae includes the bulk of modern spiders. Seven spider families currently contain more than 1000 species—all are entelegyne.

During a recent 39-year period, an average of 314 new species were described per year (12,200 total), but an annual average of 104 old names were synonymized, for a net gain of approximately 200 species per year (8800 total). Estimates of total spider diversity range from 76 to 170,000. Lower estimates mostly extrapolate from the proportions of new versus known species in taxonomic publications or are based on comparisons to well-known groups. The higher estimates take into account that many regions, particularly those richest in spider species, are disproportionately undercollected. In any case, the real diversity of spiders will never be known because a potentially great fraction will certainly go extinct before being discovered, much less described.

1. Major Lineages

a. Mesothelae

Liphistiidae is the only extant family and is limited to areas of Southeast Asia and Japan. Only a few dozen species are known, but some are common where they occur. Liphistiids retain many primitive morphological features, such as eight (rather than six or fewer) segmented spinnerets that insert anteriorly rather than terminally on the semi-segmented abdomen. Their biology may likewise represent the ecological "ground plan" for spiders. Liphistiids live in silk-lined tubes equipped with rudimentary trap doors in banks and cave entrances. Sometimes, silk "trip lines" lead away from the burrow entrance to extend the sensory radius of the animal. They are nocturnal, ambush predators. They live for 5–8 years, are remarkably sedentary, and consume a catholic diet of mainly walking prey.

b. Mygalomorphae

Mygalomorphs include the tarantulas or baboon spiders (Theraphosidae), trap-door spiders (Ctenizidae, Actinopodidae, Migidae, etc.), purse web spiders (Atypidae), funnel web spiders (Hexathelidae), and several other families with no common name. Mygalomorphs number 15 families and approximately 2200 species, but several of the families are para- or polyphyletic. A more realistic estimate is 20–30 "family-level" groups. Like mesotheles, mygalomorphs tend to live in burrows and forage at the burrow entrance or for a very limited distance around it. Some theraphosids are arboreal and spin elaborate silken retreats. Diplurids make extensive webs but are virtually unique among mygalomorphs. The venomous Australian funnel web spiders (*Atrax and Hadronyche*: Hexathelidae) were responsible for many deaths until an antivenin was developed in the 1980s. The large theraphosid baboon spiders are not seriously venomous to humans, despite their popular reputation.

c. Araneomorphae

Araneomorphs include approximately 94% of known spider species. Even the most primitive araneomorphs are very different from mygalomorphs and mesotheles. Basal araneomorphs tend to be much smaller, and most are obligate web spinners with elaborate spinnerets capable of making adhesive "cribellate" silk. Cribellate silk is adhesive due to the extremely fine threads drawn from the cribellum, the very modified and fused anterior pair of spinnerets. Adhesive silk makes feasible a greater variety of web architectures, and these basal araneomorphs spin elaborate catching webs. The sister group of all remaining araneomorphs is the family Hypochilidae; its dozen or so species are limited to the Appalachian Mountains, a few places in western North America, and equally restricted sites in China.

*i. **Haplogynae*** Haplogynae comprises 17 families of spiders of diverse habits and worldwide distribution. Filistatidae are cribellate web-spinning spiders. Their web architecture is not much different from that of the mygalomorph diplurids—a sheet that narrows to a silk-lined retreat in a tube or crevice. Diguetids make dry silk webs. The majority of haplogynes are leaf-litter specialists and are vagabond, webless predators. The venomous brown recluse spiders (Loxosceles spp.) are haplogyne sicariids. The sister group of *Loxosceles* in southern Africa (Sicarius) has also been implicated in medically serious bites.

*ii. **Entelegynae*** The remaining 70 families of spiders are the Entelegynae. Entelegynes share many evolutionary novelties. Females have a convoluted abdominal plate protecting their gonopore, which male genitalia must navigate successfully to achieve insemination. Sperm are stored in a unique "flow-through" system so that the female reproductive tract has two apertures to the outside. Female entelegynes also make special silk used exclusively in eggsacs, although its exact role is unknown. The lateral eyes possess a canoe-shaped tapetum that in some spiders enables orientation via polarized light. Although the higher phylogeny has been worked out for many entelegyne lineages, some very large ones remain unstudied, and the relationships between entelegyne lineages are also controversial.

Lycosoidea includes 11 families of mainly hunting spiders, some with common names: lynx, wolf, fishing, or tropical wolf spiders. Lycosidae (wolf spiders) and Pisauridae (fishing spiders) are common, cosmopolitan lycosoid spiders. Lycosoids occur in all terrestrial habitats, and some are semiaquatic in their ability to run across the surface of the water or dive beneath the surface. Web spinning is rare among lycosoids; some may have regained it after evolutionary loss. Most species are vagabond predators or, occasionally, tube dwellers. Active at night, they move sporadically or wait until prey approaches and then attack with powerful front legs and chelicerae. They are built strongly and run and jump with agility. The South American ctenid *Phoneutria* is venomous to humans.

Dionycha includes 17 families of spiders with two tarsal claws and a tuft of hair rather than the more common three claws and no claw tufts. The monophyly of Dionycha is by no means certain. Dionycha are also hunting spiders and have habits similar to those of lycosoids. Crab spiders (Thomisidae) wait for insect prey in flowers. Jumping spiders (Salticidae), the largest spider family, can be very brightly colored and often prefer to jump rather than walk. Their vision is superior to any other arachnid; salticids are the only sizable spider lineage that is strictly diurnal.

Orbiculariae includes 14 families of spiders and approximately 12,000 species. Most orbicularians spin prey-catching webs, but a few groups have secondarily lost the web-spinning habit. The primitive web architecture seems to be the orb—the classic spider web of radially symmetric, stiff, dry spokes supporting a spiral of sticky silk—but more orbicularian species have lost or modified the orb architecture than retained it. Web spiders rely exclusively on webs for prey capture. Araneidae (common orb weavers), Linyphiidae (sheet weavers), and Theridiidae (cobweb weavers) are the largest families. The venomous widow spiders (*Latrodectus*) are theridiids and are distributed worldwide; several species are spread by humans and are now cosmopolitan.

III. SCORPIONES

About 1256 species of scorpions are currently known in 156 genera and 18 families (Fig. 4). Scorpions are one of the better collected arachnid groups so that huge increases in diversity are not as likely as in mites or spiders. Estimates of total diversity run as high as 7000 species. All scorpions have large, obvious pedipalps modified as pincers, both body regions are broadly joined, the distal abdomen is narrowed into a flexible tail bearing a venomous stinger at the end, and ventrally. The abdomen bears a pair of comb-like sensory appendages known as pectines. Colors vary from translucent to brown or black. Curiously, they fluoresce under ultraviolet light, a discovery that has galvanized recent field research on these animals. The longest is *Hadogenes troglodytes* at 21 cm, but *Pandinus imperator* is nearly as long and much heavier. The external morphology of recent scorpions is impressively similar to Silurian fossils. Formerly scorpions were thought to be the sister group of all other arachnids because they closely resemble the extinct marine eurypterids that are the sister group of all arachnids. Better analysis of morphological data (Fig. 2), weakly corroborated by molecular evidence, suggests that scorpions are more deeply imbedded in the arachnid clade and merely retain many primitive features. The issue is controversial.

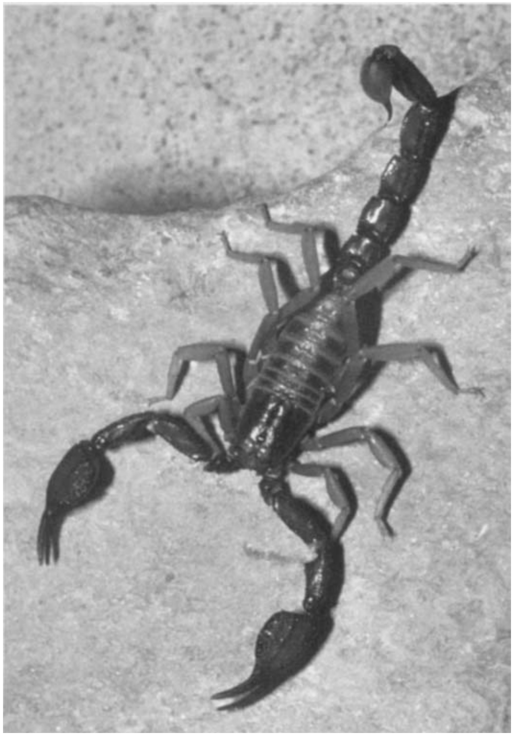


FIGURE 4 Scorpion (Scorpiones: Vaejovidae, *Uroctonus* sp.) (photograph by James C. Cokendolpher). See also color insert, this volume.

Scorpions are the only arachnids with a narrow postabdomen ("tail") terminating in a venomous sting. The sting is most often used for defense, although scorpions will sting large or strong prey. The sting of most scorpions is painful—like wasp or hornet stings—but not dangerous. Characteristically, scorpions with slender pedipalps are more prone to sting their prey, whereas those with robust pedipalps tend to crush prey. The Central American genus *Centruroides*, Brazilian *Tityus*, and Old World *Androctonus*, *Leiurus*, *Mesobuthus*, and *Parabuthus* are very venomous and medically important. In Mexico, *Centruroides* spp. sting 300,000 and kill 1000 people annually; *Androctonus*, *Leiurus*, and *Mesobuthus* kill thousands annually in Egypt and Pakistan alone. Excepting ticks that spread disease, scorpions are by far the most dangerous arachnids to humans. Scorpions are correspondingly prominent in mythology and folklore (e.g., the zodiacal constellation "Scorpio"). Scorpion venoms typically contain multiple low-molecular-weight proteinaceous neurotoxins. Scorpion blood inactivates scorpion venom, but if the venom is injected directly in a nerve, the animal rapidly dies. *Parabuthus transvaalicus* and *P. villosus* squirt venom to damage corneas, like spitting cobras.

A. Ecology

Scorpions are most diverse in deserts or similar dry areas, although they are reliably present in moist ecosystems if the temperature is not too cold. They now occur on all major landmasses except Antarctica. Favored habitats are burrows, under bark, stones, or logs, or inside small crevices. Burrows may be as deep as 40–80 cm, serving to escape the hot daytime temperatures in deserts. Because they like hard substrates and dry conditions, scorpions adapt well to human structures. Most scorpions are ground dwelling, but many species are arboreal, especially in the neotropics. In canopy fogging at four Amazonian sites in Peru, all of approximately 100 specimens were Buthidae (J. Ochoa, personal communication). A few are limited to lightless caves.

Scorpions are almost invariably nocturnal, although the African *P. villosus* is predominantly diurnal. The eyes seem to detect luminosity at best. Prey movements are detected by tarsal sense organs at distances up to 15 cm, and prey are attacked in a single motion. At distances up to 30 cm prey are located through orientation responses. The large, pincher-like pedipalps immobilize prey; thereafter, pieces are torn off by the chelicerae and digested in the pre-oral cavity before being sucked into the gut. Scorpions can be important consumers in some communities. In Israel, *Scorpio maurus* annually ate an average of 11% of the isopod population, which was not the only item in their diet. At moderate densities of 1.5 kg/ha, *Urodacus yaschenkoi* ate an annual average of 7.9 kg/ha of prey. Cannibalism and predation by other scorpion species are thought to be the most important sources of mortality, but other top invertebrate predators (e.g., spiders) and vertebrates are also important scorpion predators. Generally, mortality is highest immediately after birth, lower for intermediate-aged animals, and high for adults (e.g., 65, 30, and 60%, respectively, per year for the Australian *Urodacus manicatus*). Scorpion mortality is particularly high among males because of their high activity levels and mobility during the breeding season. Cannibalism by females is a significant cause of male death. Biased adult sex ratios of 1.2–1.4 : 1 are typical. Communal behavior, however, does occur. For example, family groups up to 15 individuals of the Brazilian *Opisthacanthus cayaporum* cooperate to construct and occupy communal chambers in the center of termite mounds. The African *Heterometrus* spp. also construct and share a communal burrow, inhabited by individuals of various ages. "Piles" of 20–30 individuals of *Centruroides exilicauda* are found in the winter months. Groups of 5 individuals of *Mesobuthus martensi*, all of the same age and all with their heads oriented toward a central spot, have been found under wet rocks in the intertidal zone.

The vast majority of scorpion species are subtropical or tropical. Point diversity (the number of species sympatric at one site) peaks in subtropical deserts and is particularly high (6 or 7, with a maximum of 12) in Baja California. Two to three species per site is more usual. *Vaejovis littoralis* reaches unusually high densities (8–12/m^2) in the drift line along the Gulf of California. The North American *Paruroctonus boreus* occurs as far north as British Columbia and Alberta, and the European *Euscorpius germanus* reaches the southern Alps. Even tropical scorpions sometimes inhabit extreme conditions; *Orobothriurus crassimanus* was collected at 5560 m in the Peruvian Andes.

B. Reproduction

Reproduction in scorpions is via a spermatophore attached to the substrate. The male completes production of the spermatophore inside his body, deposits the sperm inside, and attaches the spermatophore to the substrate, all the while holding on to the female during preliminary courtship. The spermatophore is "spring-loaded" and catapults the sperm mass into the female gonopore when a lever is touched. Scorpions are exclusively ovoviviparous or viviparous. The 1–105 young are born live and cling to the mother for the first few molts. A few species are parthenogenetic. Scorpions live 4 or 5 years (rarely 8); they do not molt as adults.

C. Phylogeny and Taxonomy

The higher classification of scorpions has changed dramatically as classical data have been reinterpreted phylogenetically. The old system proposed a few, huge, polyphyletic families about which nothing much in general could be said; now 16 or 18 families with increasingly coherent biologies are recognized. Species limits in scorpions are often difficult because scorpion genitalia are usually not species-specific. By tradition, scorpion taxonomists use the subspecies category more than most arachnologists. About 150 subspecies are recognized in addition to the 1260 species, but because these are easily distinguished they are probably distinct species. A classic example is the 25 non-overlapping, fully distinct subspecies of *Scorpio maurus*. Species-level taxonomic characters include the surface sculpturing of the exoskeleton, morphometric data, the number and position of pedipalpal trichobothria, and the hemispermatophores—internal male structures that produce the spermatophore.

1. Major Lineages

The basal division in Scorpiones is between the buthoids and remaining scorpions. New and Old World buthids are also distinct lineages. Scorpionoids and the vaejovoid–chactoid lineage are the remaining major scorpion lineages. Chaerilidae and Pseudochactidae (*Chaerilus*, 21 species; *Pseudochactus ovchinnikovi*, from Kazakhstan) are monogeneric and enigmatic; they are like none of the other scorpion families and their relation to other major lineages is obscure. They may be basal buthoid groups.

Buthidae is the largest and most widely distributed scorpion family with approximately 74 genera and 531 species. Buthidae is most diverse in the African tropics and Palearctic regions. Buthids tend to have slender, elongate pincers, a robust tail, and usually a tubercle under the sting. All scorpions considered dangerous to humans are buthids. Buthids are also the most diverse ecologically and occupy humid, mesic, and dry habitats. The small family Microcharmidae is an Afrotropical buthid segregate with two genera and six species.

The Scorpionoidea (36 genera and 355 species) is a large, monophyletic lineage that includes Bothriuridae, Diplocentridae, Heteroscorpionidae, Hemiscorpiidae, Ischnuridae, Scorpionidae, and Urodacidae. Scorpionidae lack a tubercle under the sting and the sides of the sternum are parallel. The family contains the genus *Scorpio* from the Mediterranean and Near East, much mentioned in classical Greek, Egyptian, and Christian myths, and *Pandinus*, the giant African black scorpion. The longest and heaviest scorpions are scorpionids, which are exclusively Old World.

The relationships and monophyly of the vaejovoid–chactoid lineages (43 genera and 495 species: Chactidae, Euscorpiidae, Iuridae, Scorpiopidae, Superstitionidae, Troglotayoscidae, and Vaejovidae) are the most problematic areas of scorpion higher taxonomy and phylogeny. The large families Iuridae and Chactidae, in particular, are doubtfully monophyletic, although each includes many clearly valid groups. Together, the vaejovoid–chactoid lineages comprise about one-fourth of known scorpion species, including the most common species in North America.

IV. OPILIONES

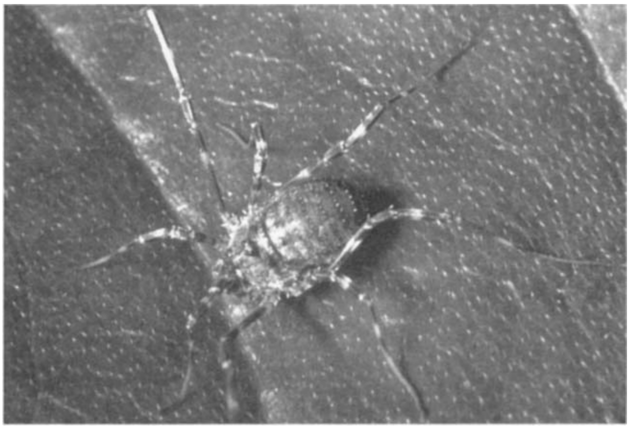


FIGURE 5 Harvestman (Opiliones: Phalangiidae, *Hadrobunus* sp.) (photograph by Jonathan A. Coddington).

The described world fauna of Opiliones (harvestmen or daddy long legs) comprises approximately 44 families, 1554 genera, and about 5000 species (Fig. 5). The largest harvestman is *Trogulus torosus* at 2.2 cm long. The anterior and posterior body regions are broadly joined and the abdomen is rather short, giving the body a wider and rounder appearance than those of other arachnids. The second pair of walking legs is usually the longest. Opilions have just two eyes (cave or litter groups are sometimes blind). All harvestmen have a pair of glands that open via large pores on the anterolateral margins of the body; the function of their secretions is apparently diverse. Harvestmen are the only arachnids in which males have a penis. Females have a long, flexible, and extensible ovipositor (as do many mites). The most common group in north temperate regions, the Phalangioidea (daddy long legs), have soft and flexible bodies, weak mouthparts, and extremely long legs (commonly 15 times the body length), but just as many harvestmen are fantastically armored with bizarre, huge chelicerae, raptorial pedipalps, and short, stiff legs. Others are mite-like, and still others are dorsoventrally flattened. Colors are usually subtle patterns of brown, gray, or black, but tropical forms can rival anything seen in spiders or mites. In the long-legged forms, the distal leg tip is divided into numerous false segments that form a prehensile tip. It can be wrapped around objects to achieve a very firm and adaptable grip. The long second legs may double as feelers.

The respiratory system is exclusively tracheate. The very long-legged Phalangioidea have accessory spiracles on distal leg articles. Touch and vibration perception, as in most other arachnids, seems to be the dominant sense. The eyes at best distinguish light and dark and direction of light. Opilions consume a broader diet than any other arachnids other than mites. The basic pattern is predation, but some, for example, specialize on snails, which are otherwise rarely consumed by nonacarine arachnids. Opilions also are known to eat dead insects, fruit, and decaying vegetable matter. Unlike other arachnids, harvestmen can ingest solid particles, as shown by sclerotized bits in their excrement.

Many harvestmen defend themselves against attack by shedding legs. One study of two species in Louisiana found that about half the animals had lost one leg, but less than 10% lacked three. Seven-legged harvestmen seem to survive and function as well as intact animals. A shed leg continues to jerk and twitch attractively for minutes, permitting the harvestman to escape. All harvestmen have paired "repugnatorial" glands on the front margin of the body. When legs are pinched a droplet of fluid appears at the orifice which may be dabbed on an attacker with a leg or allowed to evaporate. The secretions can also help aggregating species to find each other, and some show broad antibiotic and antifungal activity, presumably useful to litter and soil-dwelling forms. Quinones are a major ingredient. Soil-dwelling harvestmen and the short-legged Laniatores are slow compared to long-legged harvestmen.

A. Ecology

In general, harvestmen prefer moist, or at least not xeric, environments. The mite-like Cyphophthalmi live in dark leaf litter, caves, or under stones. The largely tropical and usually short-legged Laniatores move slowly over vegetation or the forest floor. The usually long-legged Eupnoi can be anywhere, but their very long legs with prehensile tips are specialized for crossing the large gaps between the leaves of trees, shrubs, and herbs. The northern European opilionid fauna comprises approximately 24 species and that of North America approximately 235 species. In temperate regions, diversities of more than a dozen species per hectare are uncommon.

B. Reproduction

Uniquely among non-acarine arachnids, sperm transfer occurs via the male penis. Mating occurs quickly and seemingly without courtship. The male faces the female and pushes the penis underneath her body or between her chelicerae and into the gonopore. After insemination, the sexes separate and continue their solitary wanderings. Females use their long, flexible ovipositors to deposit eggs in suitably protected areas. Trogulids deposit their eggs only in empty snail shells, and other groups oviposit beneath stones, deep into soil, underneath bark, or in bore holes in plant stems left by insects, usually abandoning the eggs once laid. Newly hatched animals are active and resemble adults. Five to eight molts to maturity are common. An unusual reversal of sex role occurs in the Panamanian harvestman *Zygopachylus albomarginis*. Males fight to occupy existing mud nests or construct their own. Females wander between nests, courting the males, mating, and ovipositing in a series of nests. They have nothing more to do with the offspring. Males accumulate eggs of different ages and from different females and defend the eggs against conspecifics and ants.

C. Phylogeny and Taxonomy

The phylogeny of Opiliones has recently been clarified at the superfamily level, but many additional changes are expected in familial arrangements. Some families seem to be based only on primitive features (e.g., Travuniidae, Phalangodidae, and Triaenonychidae). Eupnoi and Dyspnoi classically formed the suborder Palpatores, but increasing evidence indicates that this taxon is paraphyletic. The number of recognized families has approximately doubled in the past 20 years. At the species level the morphology of male genitalia is especially diagnostic.

1. Major Lineages

a. Cyphophthalmi

This suborder contains five (or six) families and about 100 species. It is sister to all remaining Opiliones (the "Phalangida"). Cyphophthalmi are eyeless, live in deep moist leaf litter or caves, and range in size from 1 to 7 mm. The animals have a hardened plate covering the entire dorsal surface, and they resemble mites. *Siro* is North American and European, but the family also occurs in Southeast Asia, Turkey, Japan, Mexico, and South Africa. Life spans of up to 7 years have been reported.

b. Eupnoi

This group contains two superfamilies, including the classic daddy long legs (Phalangioidea: Phalangiidae, Sclerosomatidae, Megalopsalididae, and Neopilionidae) of soft-bodied, long-legged harvestmen. *Phalangium opilio* is common around buildings and introduced throughout the world. The sclerosomatid *Leiobunum* spp. are common in North American and European forests, in which they move easily across the upper vegetation. They are predators and scavengers. Caddoidea includes only one or two families of harvestmen with enlarged eyes and short legs, sometimes common on tree trunks.

c. Dyspnoi

This group also includes two superfamilies, but has many fewer species than Eupnoi. Ischyropsalidoidea contains three families and just seven genera. *Ischyropsalis* is European and feeds on snails. Troguloidea contains four families, two monogeneric and several genera in the remaining families. *Nemostoma* is common in caves. Trogulidae are peculiar harvestmen that look like giant, flattened mites. Legs are very short. Trogulids live under stones and in leaf litter; snails are an important part of their diet.

d. Laniatores

This group is morphologically more diverse than the previous groups and tends to be more diverse in the tropics, including many colorful species. Three superfamilies are recognized: Traviunoidea (4 families), Oncopodoidea (1 family), and Gonyleptoidea (18–20 families). Colloquially, laniatorids are known as "short-legged" harvestmen because the most common laniatorids do have short legs, but some agoristenids, cosmetids, gonyleptids, caelopygines, progonyleptoidellines, and mitobatines have legs comparable in length to those of Palpatores. The Gonyleptoidea contains 18–20 families with raptorial pedipalps for prey capture and enlarged fourth coxae with spreading appendages—perhaps a defense against being swallowed whole by predators or dragged down the burrow of a parasitoid. Gonyleptidae are typical heavily armored and spiny, often colorful, and usually larger than 4 mm. They are exclusively Neotropical and, with more than 100 genera, are one of the largest families. The closely related Cosmetidae superficially resemble gonyleptids—large, heavily armored, often colorful, and slow-moving New World harvestmen. Phalangodidae is a large cosmopolitan family of more than 150 genera defined mainly by the features they lack. For years it has been used to file taxa with no obvious relatives, and therefore its biology makes little sense. Although most diverse in the northern neotropics, their distribution includes southern North America. Their pedipalps are flattened with spiny margins.

V. SMALLER ARACHNID ORDERS

Amblypygi (whip spiders or tailless whip scorpions), Uropygi (whip scorpions or vinegaroons), Schizomida (no common name), and Palpigradi (micro-whip scorpions) are all small, unfamiliar arachnid groups that are closely related to each other and to spiders (Fig. 2). Like primitive spiders, all have two pairs of abdominal book lungs, although the second pair is missing in the tiny Schizomida and palpigrades lack both. The first pair of walking legs is elongate and feeler-like, with false articulations in the terminal articles to promote flexibility. Pseudoscorpiones (pseudoscorpions), Solifugae (wind spiders), and Ricinulei (no common name) are a miscellany of remaining arachnid orders related to various orders already discussed, as illustrated in Fig. 2.

A. Amblypygi

Approximately 125 species of amblypygids in 20 genera and 5 families are known (Fig. 6). The American *Acanthophrynus coronatus* is the biggest, at about 4.5 cm in body length. The usual adult body size is about 4–6 cm. The first walking legs of amblypygids are enormously long, and fully stretched a large animal can span 50 cm. Whip spiders have no tail, and their pedipalps are modified into fierce-looking, spiny raptorial appendages. Amblypygids are easily recognized by their extremely long front legs and raptorial pedipalps. Colors are dull brown or black. The body is flat and leg insertion twisted so that limbs fold in the same plane as the body (like a crab), permitting the animals to edge through thin openings such as cracks in hollow trees. They move sideways more easily than forward or back. Amblypygids hunt by drifting their long front legs gently over the surface around them to locate prey and using their raptorial pedipalps to pounce. Like spiders, all the abdominal ganglia have migrated into the prosoma in which, fused, they form a brain. Reproduction is via a spermatophore. Like uropygids, amblypygid females carry their egg clutches inside a membrane of dried mucus glued to their ventral abdomens. The 15–50 young hatch and remain inside this membrane until they have undergone their first molt. The young cling to the mother for a short time.

FIGURE 6 Whip spider or tailless whip scorpion (Amblypygi: Phrynidae, *Phrynus sp.*) (photograph by Jonathan A. Coddington). See also color insert, this volume.

Whip spiders live in subtropical and tropical areas, in forests and often in caves. They are exclusively nocturnal and fairly common. During the day they hide in hollow trees or logs, under loose bark, or under large logs. Only one species lives in leaf litter and is not known to burrow. Their diet seems to be a broad range of smaller arthropods in their environment.

B. Uropygi

About 100 species of whip scorpions are known in 16 genera and one family, Thelyphonidae (Fig. 7). At 7.5 cm, the largest is *Mastigoproctus giganteus* of North America, but 3–5 cm is usual. Whip scorpions are easy to recognize by the long posterior whip or flagellum (highly modified terminal abdominal segments). Colors are brown to black. Uropygids have defensive anal glands that accurately spray an acid-smelling fluid at attackers. The smell explains the common name "vinegaroon." The fluid of *M. giganteus* is 85% acetic acid but it also contains substances to reduce the surface tension of the epicuticle so that the acetic acid can spread widely and penetrate. Vinegaroons are supposedly not sensitive to their own spray. Reproductive habits are known only for a few species, but presumably sperm transfer occurs via a spermatophore glued to the substrate in all species. Some species have a lengthy courtship—10 hr to several days. Females keep their 12–40 eggs attached to their ventral abdomen. Female *Thelyphonus* build a deep burrow and do not feed while guarding the eggs for 4 or 5 weeks. Uropygids may live 6–8 years or even longer.

Uropygids are tropical to subtropical animals. They hide in leaf litter, burrows, under logs and rocks, and inside dark crevices or holes during the day, emerging at night to hunt. Not much is known about their prey, but presumably it consists of other small ground-dwelling insects, arachnids, and crustaceans, which they crush between their pedipalpal segments prior to ingestion.

FIGURE 7 Whip scorpion (Uropygi: Thelyphonidae, *Mastigoproctus* sp.) (photograph by Jonathan A. Coddington). See also color insert, this volume.

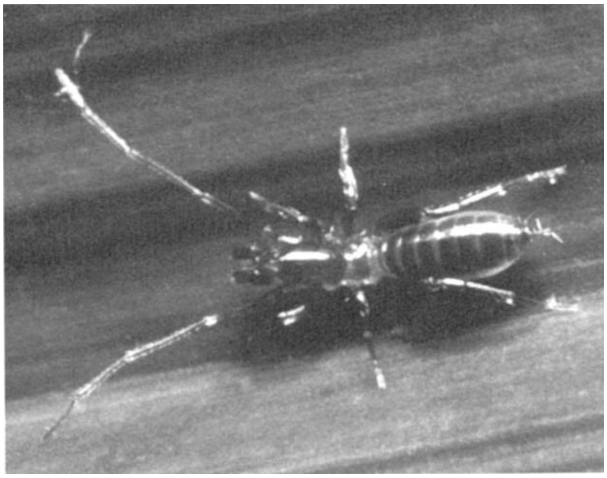


FIGURE 8 Schizomid (Schizomida: Hubbardiidae) (photograph by Jeremy A. Miller).

C. Schizomida

About 200 species in 30 genera and two families (Protoschizomidae and Hubbardiidae) are known (Fig. 8). Schizomids (there is no common name) are most like tiny uropygids, but the abdominal flagellum is short (three segments but rarely five). Colors are usually light yellow to tan to dull green. Assiduous searching in moist tropical litter usually turns up a schizomid. The largest schizomid known is *Agastoschizomus lucifer* at about 12.7 mm long, but 3 mm is typical. Most schizomids lack eyes entirely and live in moist tropical or subtropical leaf litter, under stones, in logs, moist crevices, caves, and so on. Reproduction is via a spermatophore attached to the substrate, after which females lay 6–30 eggs that are glued to the ventral female abdomen until the young emerge. At two Amazon sites schizomid abundances ranged from 5 to 110 animals per square meter per month.

Like the uropygids they have been reported to produce a defensive chemical smell. They can move backwards rapidly and have enlarged femora apparently used to hop backwards.

D. Palpigradi

There is only one family (Eukoeneniidae) with five or six genera and about 80 species. Palpigrades (micro-whip scorpions) are tiny, light yellow to white, delicate soil and leaf litter specialists with a cosmopolitan distribution. The largest is *Eukoenenia draco* at 2.8 mm long, but 1 or 2 mm is typical. They resemble the young of Uropygi. Like whip scorpions, they have a wide preabdomen and a multi-segmented whip-like postabdomen. Like many soil organisms, palpigrades lack most of the organ systems required by large animals that live in drier and less constant environments. They lack eyes, respiratory organs, and a circulatory system but have innervated setae that detect vibrations. Most species are known from the tropics; several palpigrades apparently live in intertidal or shallow marine habitats. *Eukoenenia janetscheki* was the only species found at two Amazon sites, but it was fairly abundant (30–120 animals per square meter per month).

E. Ricinulei

One family (Ricinoididae), three genera, and 53 species of ricinuleids are known (Fig. 9). Ricinuleids, like schizomids and palpigrades, are soil-dwelling specialists. The largest is *Ricinoides afzelli* at 10 mm. long, but most are 3–5 mm. Unlike the preceding forms, ricinuleids are heavily armored with a thick exoskeleton and a kind of visor or flap (the cucullus) that folds over the mouthparts. Colors are reddish to brown or black. Males have third legs modified for sperm transfer. Some have "eye-spots" that are able to sense light. Females lay one or two eggs at a time, which they may carry about with them. Adults may live 10 years. Formerly thought to be extremely rare, ricinuleids are not uncommon in Neotropical leaf litter. They are also known from Africa, but not from the Asian tropics.

F. Pseudoscorpiones

About 3100 species of Pseudoscorpions (or "false scorpions") have been described, classified in approximately 430 genera, 24 families, and two suborders (Epiocheirata and Iocheirata) (Fig. 10). The North American fauna comprises about 350 species. One local fauna in the Amazon comprised nine species and Mediterranean ecosystems around Perth, Australia, and other Old World tropical moist ecosystems are comparable (M. Harvey, personal communication). Pseudoscorpions look like tiny scorpions without the tail and pectines. Colors are tan to dark brown or black. Their chelate pedipalps can have venom glands and the chelicerae have silk glands. The largest is *Garypus titaneus* at 12 mm long, but most are 2 or 3 mm. They are flattened for life in crevices, usually under bark or in leaf litter, but a few forms live on ocean shores where they are truly intertidal. Numerous species live among the hairs of mammals where they feed upon other arthropod parasites. At least one species has the body, legs, and palps modified for living on a Celebese rat. *Chelifer cancroides* is cosmopolitan and lives in houses where it hunts book lice and other small insects. Pseudoscorpions are sometimes found in aggregations of dozens of individuals. Reproduction, as in most other arachnids, is via a spermatophore. The female builds a silken nest

FIGURE 9 Ricinuleid (Ricinoididae) (photograph by Jonathan A. Coddington). See also color insert, this volume.

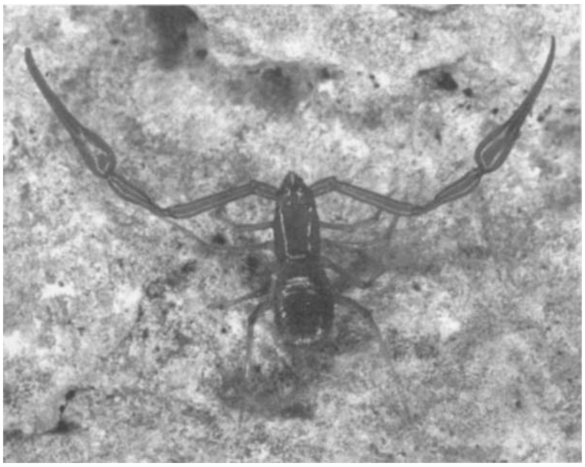


FIGURE 10 Pseudoscorpion (Pseudoscorpiones: Neobisiidae, *Tartarocreagis infernalis* (Muchmore) (photograph by James C. Cokendolpher).

and secretes a "brood sac" attached to her body and in which she nourishes the 3–40 embryos with maternal secretions. Juveniles leave the mother soon after hatching. Pseudoscorpions are famous for their tendency to disperse by phoresy. Mature females (rarely males or juveniles) use their pedipalps to hold onto larger insects, which carry them from place to place as they fly about. Phoretic specimens in fossil amber show that this behavior is at least 10 million years old.

G. Solifugae

About 1065 species of solifuges are known, grouped in about 153 genera and 12 families (Fig. 11). The greatest known diversity occurs in Namibia. Wind spiders (sun spiders or camel spiders) look somewhat like hairy, stout, fast-running spiders, but their chelicerae are fearsomely enlarged and both the leg-like pedipalps and the long and slender first legs are used as feelers. The fourth femurs bear ventral sensory structures shaped like inverted mushrooms. Colors are tan to sandy yellow. The largest is *Galeodes caspius* at 7 cm, but 2 or 3 cm is usual. Wind spiders can run extremely fast (53 cm/sec) for short bursts, but like most arachnids they cannot sustain rapid locomotion for long periods. The solifuge respiratory system is only tracheate and is comparable in complexity to that of insects. They are generally nocturnal and exclusively carnivores and, like any of the larger arachnids, will take small mice, lizards, amphibians, and so on if the opportunity arises. A few species feed only on termites. Unlike most other arachnids, solifuges are active, cursorial predators. Wind spiders are most common in tropical and subtropical dry or desert regions, but they are absent from Australia. Many inhabit burrows, where they may stay up to 9 months of the year, depending on rainfall. Reproduction is via spermatophore, which the male produces during courtship and transfers into the female gonopore with his chelicerae. The female digs a deep burrow to deposit egg masses that may contain 5–164 eggs. Females may lay 1–5 egg masses. Some females remain with the eggs until hatching. Newly hatched wind spiders are gregarious and remain in the burrow for the first two molts.

VI. ACARI

Mites and ticks are classified as Acari, from a Greek word meaning a thing too tiny to be divided (Fig. 12). At more than 45,000 species, Acari is already the most diverse arachnid order, but acarologists estimate that between half a million and more than 1 million species actually exist. Despite being far more abundant and diverse than any other order of arachnids, mites and ticks are nonetheless much less familiar to most people than spiders, harvestmen, or scorpions. Likewise, scientific knowledge of mites lags far behind our understanding of most other arachnid groups. The primary reason for both public and scientific ignorance of mites is their small size.

Since we sometimes find them on ourselves and our pets, because they are relatively large, and because of their role as serious vectors of human disease, the ticks are probably the most familiar group of mites. In fact, ticks (which have their own name only because of their

FIGURE 11 Wind spider or solifuge (Solifuga: Eremobatidae) (photograph by Jonathan A. Coddington). See also color insert, this volume.

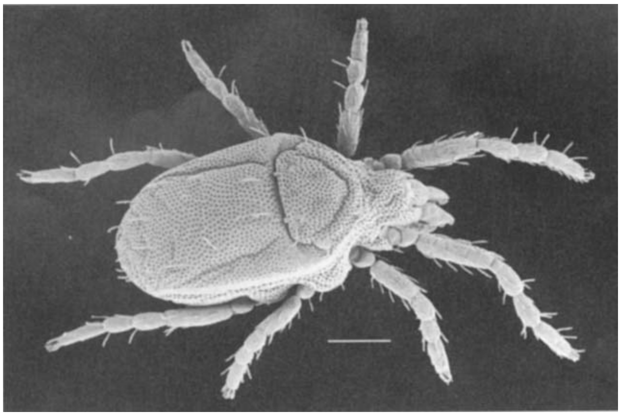


FIGURE 12 Free-living yellow moss mite (Acari: Penthalodidae, *Stereotydeus* sp.) (photomicrograph by David Evans Walter; scale bar = 100 μm).

importance to humans) represent a small offshoot on the evolutionary tree of mites. Other kinds of mites are vastly more common and diverse than the ticks and include not only other parasites but also followers of many other ways of life. In this article, unless otherwise distinguished, the term *mites* is intended to include ticks and will be used interchangeably with Acari.

The Acari are distinguished from other arachnids principally by the lack of conspicuous body segmentation, the broad union of the leg-bearing part of the body (the podosoma) with the posterior part (the opisthosoma) to form a single unit (idiosoma), the aggregation of the mouthparts into a distinct anterior body region (the gnathosoma), and the first active immature stage, the larva, having only three pairs of legs. The gnathosoma is not a true head, like that of an insect, in that it possesses neither eyes nor antennae and does not contain the brain. (Most mites have no eyes at all, but those that do bear them on the idiosoma. The brain is situated anterior to the stomach in the idiosoma.) Although usually compact in form, like that of a spider mite or tick, the overall shape of the mite body varies greatly. Species that live in polypore fungi or hair follicles are elongated to fit their habitat, and mites found deep in soil tend to be worm-like. Ticks and some other mites can tolerate enormous distention of the body wall to accommodate food.

Like other chelicerate arthropods, mites have no antennae, but in many groups the first pair of legs carry chemosensory and tactile organs and are not used, or little used, for locomotion. The pedipalps, integrated laterally into the gnathosoma, are also primarily sensory in function. The chelicerae, the primary food-getting organs, vary greatly in form, in accord with the varied feeding habits of mites. In many species, the chelicerae of males are so modified for sperm transfer that adult males scarcely feed.

A. Ecology

Mites are found in virtually every habitat on Earth. On land, legions of free-living mites populate soils in all habitats, often numbering in the hundreds of thousands per square meter in the soil (50,000–250,000 per square meter, down to 5 cm depth, in forests; 20,000–100,000 in grasslands; and 500–1000 in deserts). In many soils, including deserts, mites can be found at depths up to 10 m, where they follow plant root systems. Although the habits and behavior of many soil mites remain a mystery, it is known that soil mites fill many ecological roles. Many eat decaying organic matter (and the microorganisms it contains); others consume fungi, algae, and bacteria; and still others are predators, feeding on nematodes, other mites, or various stages of insects. Oribatid mites in the soil facilitate plant growth by dispersing the spores of mycorrhizal fungi, which form mutually beneficial and often essential associations with higher plants.

Mites are common and often abundant on plants, on which many species feed on leaf, stem, or root tissue by piercing cell walls. Others specialize on plant reproductive parts, including nectar, pollen, and fruit. Still other plant-dwelling mites are fungivores (eating fungi on leaves or stems) or predators, mostly of plant-feeding mites. Unlike life in the soil, life on plant surfaces constantly threatens mites with desiccation. Hiding from direct sun on the underside of leaves offers some protection because the "boundary" layer of still air is kept relatively moist by plant transpiration.

In addition, many plants produce domatia—hollow, mite-sized cavities or pockets in the leaf. The adaptive function of domatia is controversial, but experiments have shown that removal of domatia from some plants that normally have them reduces the density of beneficial predatory mites and increases the density of herbivorous mites.

Some fungus-feeding mites have specialized cavities or pouches in the body wall (sporothecae) in which they carry spores or propagules of fungi that they depend on for nourishment.

Parasitic and phoretic mites exploit most terrestrial vertebrate species, most other arthropod groups, mollusks, and even marine invertebrates (Fig. 13). Most species of birds, for example, generally support many "feather" mite species found nowhere else. Four or five different, host-specific species inhabit different feather

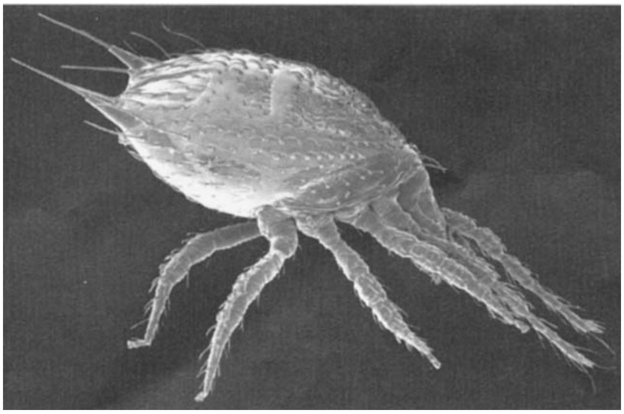


FIGURE 13 Parasitic sejine mite (Acari: Sejiidae, *Epicroseius* sp.) (photomicrograph by David Evans Walter).

types (flight feathers, tail feathers, or body feathers), whereas different species live on the surfaces of feathers, others in the interior lumen of the quills, others on or in the skin, and yet others in the nasal passages. One particular species of parrot hosts more than 20 species of feather mites. Bird nests support additional, phoretic species that feed on debris or on other nest associates. Both parasitic and phoretic species generally colonize young birds while in the natal nest and travel with the host until it builds its own nest. Similar sets of species inhabit the bodies of mammals and reptiles and, to a lesser degree, amphibians.

Mites are also common associates of insects and other invertebrates. Some are parasitic and deleterious to their hosts, such as the tracheal mite and *Varroa* mite parasites of honeybees or moth ear mites. Some are commensal (e.g., feeding on insect waste products). Others are believed to benefit the host, perhaps by consuming other harmful mites, parasitic nematodes, or decaying matter that would otherwise nourish fungi or bacteria that might threaten the host or its young. Some insects have acarinaria, special pockets in the outer body wall, that are routinely occupied by phoretic mites. The very existence of acarinaria, which have no other known function other than to house and transport mites, is evidence for a net beneficial relationship between these mites and their insect hosts. Parasitic mites have been shown to be capable of mediating horizontal gene transfer between insect host species.

All ticks require blood from a vertebrate host for development and reproduction but live off the host when not feeding or mating. Chiggers and many other mite species must live a portion of their lives, and others all their lives, as parasites of animals. Other mites exploit animal hosts not for food but strictly for transport—a phenomenon called phoresy (see the following section). Parasitism is thought to have evolved, in some lineages, from phoretic associations.

Mites are common and diverse inhabitants of freshwater and marine habitats. In fresh water, adults and nymphs prey on other mites, crustaceans, and aquatic insects. The larvae of aquatic mites are commonly parasitic on insects and crustaceans, and nymphs and adults of unionicolid mites parasitize bivalve mollusks. Other aquatic mites graze algae and fungi growing on aquatic plants. Mites are common on seashores, including the intertidal zone. The sea is home to predatory and algae-feeding mites. Some mites have been found at depths up to 7000 m at hydrothermal vents, and one group lives in the gut of sea cucumbers. Mites are the only arachnids found in the ocean depths.

Closer to home, even the cleanest office, classroom, or home normally supports dust mites (e.g., *Dermatophagoides* spp.), which live on organic matter in house dust. Some mites are stored-product pests, feeding on grains, cheese, and other dry foods. Even more intimately, as you read this encyclopedia there is a very good chance that the follicles and sebaceous glands of your forehead or eyelids house normally benign, microscopic mites (*Demodex* spp).

B. Reproduction and Dispersal

Reproductive patterns in mites are remarkably diverse. Many mites are ordinary diploids (having an equal number of chromosomes from each parent) with two sexes, requiring fertilization by the male for females to reproduce. In other groups, males are absent, and females bear female young from unfertilized eggs. In still other mites (e.g., spider mites), females require sperm from a male to produce (diploid) daughters, but sons are haploid (they have only the maternal chromosomes) since they are produced from unfertilized eggs. In other species (e.g., some phytoseiid mites), both sexes are diploid and all eggs require fertilization to develop, but only the mother's chromosomes are active in males (called paternal genome loss).

Modes of sperm transfer are diverse. Some mites transfer sperm directly through copulation, using legs, palps, or other specialized body parts to accomplish the transfer. Other mites rely on spermatophores placed on the substrate, with or without active enticement of the female to take up the sperm.

Development of the egg takes place in most mites after oviposition. In some species (e.g., moth ear mites and hummingbird flower mites) the eggs are laid *communally*. In other mites, the young leave the mother's body as active larvae. In some mites (e.g., *Siteroptes graminum*) sisters and brothers both hatch and mate within the mother's body, and both mother and sons die when the fertilized daughters emerge—truly material for a Greek tragedy.

The development of most mites includes one or more active stages between egg and adult. (Some Heterostigmata have no active immature stages.) In nearly all groups, the first active immature stage, called the larva, has three pairs of legs, and later stages, including the adult, have four pairs. (The plant-parasitic Eriophyoidea are exceptional in having two pairs of legs in all active stages, and females of many species of insect-parasitic Podapolipidae have three, two, or even only one—the anterior—pair of legs.) The number of nymphal stages between larva and adult varies from none to three, depending on the group. Several groups of mites have metabolically inactive stages that can endure stressful periods. Life expectancy among mites also varies widely. Some mites live only about 1 week or so (e.g., hummingbird flower mites), whereas others require several years to complete the life cycle (e.g., certain ticks, oribatid mites, and water mites).

Because mites are so small and (unlike most insects) are wingless, selection for effective dispersal from one food patch to another has produced many special adaptations. Parasitic mites, for example, must have a way to find a new host—or ensure that their offspring do—before the current host dies or becomes unsuitable. In the case of plant-feeding mites, some species disperse aerially, either by simple passive transport in the wind or, in the case of some spider mites (Family Tetranychidae), by "ballooning" on strands of silk. (The silk-producing glands in these mites are not homologous to those of spiders.) Other plant-feeding mites disperse phoretically on plant-visiting animals. For example, nectar- and/or pollen-feeding insects and birds on all continents carry nectar- and/or pollen-feeding mites as hitchhikers—a way of life that has many independent evolutionary origins.

Some groups of mites have special, non-feeding stages adapted for phoretic dispersal. In others, adults and sometimes later nymphal stages mount phoretic hosts to reach new hosts or habitats when local mite population density becomes high, when the host is ready to disperse (e.g., emerging adult insects), or when a food plant becomes unsuitable.

C. Phylogeny, Taxonomy, and Current Knowledge

Only 90 species of Acari were known to Linnaeus, all of which he placed in the genus *Acarus*. Currently, perhaps 45,000 species in 431 families and 3672 genera have been described compared to 30,000 species estimated 30 years ago. The rate of description of new species is still very high—several hundred species per year—and every systematic acarologist has dozens or hundreds of undescribed species sitting on shelves in vials of alcohol, awaiting description and classification. In fact, the taxonomic impediment to full knowledge of the Acari is monumental:— The number of practicing, modern mite systematists is no more than 50, and the rate of training of new systematists is alarmingly low. Moreover, exploration and species discovery in tropical forests (especially the forest canopy), lakes, and streams is still in its earliest stages. Most estimates of the total world fauna of Acari range from 500,000 to more than 1 million species, but some acarologists even suggest that mites rival insects in worldwide number of species. In any case, mites represent by far the most species-rich and ecologically diverse arachnid group. Table I gives summary data on mite diversity.

1. Major Lineages

Given the highly incomplete knowledge of the Acari, it is not surprising that their phylogeny, classification, and nomenclature are far from settled. It is generally recognized, however, that the Acari comprise three lineages that diverged from one another in very ancient times: the Opilioacariformes and Parastiformes (which together constitute the Anactinotrichida) and the Acariformes (which constitute the Actinotrichida). Their relationships to one another and to non-acarine arachnid groups have been debated for decades, but the current opinion appears to be that the Acari, so defined, is a monophyletic group.

The taxonomic rank of acarine lineages is also controversial. The tradition among zoologists in general and arachnologists is to treat Acari as an order, but acarologists now treat it as a subclass or even a class, with subsidiary mite taxa then being orders or superorders. Clearly, this is not the place to attempt a resolution of this controversy. The taxonomy in the following sections reflects the prevailing acarological point of view that major groups should rank as orders, despite the incongruity with the rest of this article.

a. Order Opilioacariformes

The small order Opilioacariformes (with 17 species described of an estimated 85–170 total) with a single family (Opilioacaridae) are large mites (1.5–2.3 mm in length) that somewhat resemble small harvestmen. Some genera occur in dry climates, often under litter or stones, and others in tropical forest litter, where they feed as omnivores or predators.

b. Order Parasitiformes

The Order Parasitiformes includes three Suborders: Holothyrida, Mesostigmata, and Ixodida.

*i. **Suborder Holothyrida.*** The Holothyrida is a small group (with 32 species described, of an estimated 160–320 total) of large (2–7 mm) saprophagous and predaceous mites of temperate and tropical forest litter, known only from Pacific-Indian Ocean islands and the Australasian and Neotropical regions.

*ii. **Suborder Mesostigmata.*** The cosmopolitan suborder Mesostigmata, in contrast, is extremely rich in species (with 11,632 species described of an estimated

100,000–200,000 total) and diverse in habits, ranging in size from 0.2 to 4.0 mm. Mesostigmatic mites range from saprovores (eating dead or decaying organic matter), fungus feeders, predators living in soil, litter, beach wrack, dung, or rotting wood, to pollen and nectar feeders. Repeatedly, mesostigmatic lineages have evolved close relations with other arachnids (spiders, amblypygids, and scorpions), myriapods, insects, and vertebrates and their nests. Some are endoparasitic (living inside vertebrate respiratory tracts), and others are ectoparasitic. Many others are phoretic, with a variety of feeding habits. Some prey on mite and insect pests of orchards and stored food products. Some of these beneficial predators have been genetically engineered to be more acaricide resistant, enhancing integration of biological and chemical control of pest mite species.

iii. Suborder Ixodida. With 880 species described, of an estimated 1000–1200 total, the ticks (suborder Ixodida) are taxonomically the best-known major group of Acari primarily because of their medical and economic importance. Ticks are considerably larger than most other Acari; adults range in size from 1.7 to 12.7 mm in the unfed state. All ticks feed on the blood of terrestrial vertebrates as ectoparasites. (Sea snakes also have ticks.) Ticks are capable of carrying and transmitting to humans and their domesticated animals more kinds of disease-causing organisms than any other group of blood-feeding arthropods. These agents include viruses, spirochaetes, rickettsiae, anasplasmas, bacteria, piroplasma, and filariae. Important tick-borne human diseases include Russian spring–summer encephalitis, Colorado tick fever, African relapsing fever, Lyme disease, Rocky Mountain spotted fever, Siberian tick typhus, Q-fever, monocytic and granulocytic ehrlichiosis, Kyasanur forest disease, and tularemia.

c. Order Acariformes

The order Acariformes comprises three suborders—Prostigmata, Astigmata, and Oribatida—with a fourth group, the Endeostigmata (120 described species, perhaps 1200–2400 total) of uncertain rank and affinities but often treated as Prostigmata.

i. Suborder Prostigmata. Prostigmatic mites (with 17,000 described species of an estimated 320,000–640,000 total) range in size from 0.1 to 2 mm, with a few of the giant red velvet mites as large as 16 mm, and vary widely in body form, color, habitat, and feeding adaptations. They include predatory and omnivorous species living in organic strata of soils and on algae, lichens, mosses, trees, and shrubs. The Prostigmata also include obligately plant-feeding groups, including the spider mites (Tetranychoidea), false spider mites (Tenuipalpidae), and eriophyoid mites, many of which are serious economic pests of field crops, orchards, and greenhouse plants, in some cases acting at vectors of damaging plant viruses. Some groups inhabit the soil; species of the family Nematalycidae have been found at depths of up to 3 m in sand dunes. Species of at least five families of prostigmatic mites are known from Antarctica, and others live in caves. Some prostigmatic groups live in the sea, where they are predaceous or algivorous. Many families of this order are specialized for life in springs, streams, rivers, waterfalls, lakes, bogs, or aquatic interstitial habitats of all descriptions. Some groups live in thermal waters; mites of the genus *Thermacarus* inhabit hot springs, thriving in water up to 50°C, in which the larvae parasitize amphibians. As in the Mesostigmata, parasitism has arisen repeatedly in independent lineages within the Prostigmata. Some are ectoparasites of slugs, scorpions, insects, or vertebrates; some are parasitoids of eggs or larvae of insects; and some are endoparasites of insect respiratory and reproductive organs, vertebrate respiratory tracts, the quills of birds, the skin of turtles, the mantle cavity of mollusks, the guts of echinoderms, or the gills of decapod crustaceans. A few species in the order are vectors of human disease (scrub typhus), and some cause mange and other skin diseases in domesticated animals. *Demodex* spp, which live (generally benignly) in hair follicles and sebaceous glands of the human face, are prostigmatic mites, as is the straw itch mite *Pyemotes tritici*, which can cause severe skin lesions in humans.

ii. Suborder Oribatida. The Oribatida is remarkable in being the only major group of mites in which a great diversity of species (11,000 described species of an estimated 33,000–110,000 total) has evolved without the evolution of parasitism. Although a few species are phoretic as adults on insects, the rest of this vast radiation are free-living inhabitants of the soil, forest litter, tree holes, bark, twigs, leaves, mosses, lichens, algae, freshwater vegetation, bogs, and the intertidal zone. Most oribatid mites feed on fungi, but some consume dead woody material or algae, and a few are predators on nematodes, rotifers, and other small invertebrates. Oribatida (and many species classified as Astigmatata, which are actually derived oribatids) feed on particulate matter. Through their feeding, they help to maintain soil structure. Unknown for other mites, many oribatid species sequester calcium and other minerals in their thickened cuticle. As with Prostigmata, oribatid species are also found on continental Antarctic.

Some oribatids are intermediate hosts of cestode tapeworms whose final hosts are herbivorous mammals. Oribatid mites range in size from 0.2 to 1.5 mm.

*iii. **Suborder Astigmata.*** The Astigmata, closely allied to the Oribatida, are primarily associates of arthropods, vertebrates, and other animal groups, although a few species are free-living in all life stages (with 4500 described species of an estimated 90,000–180,000 total). This lineage is thought to be derived from within the Oribatida, but a revised classification, with names of major groupings to reflect this relationship, has yet to be promulgated. A key evolutionary innovation of the Astigmata is a nonfeeding, immature stage especially adapted for phoretic dispersal or tolerance of adverse conditions. Astigmatic mites include fungus feeders, a few plant feeders, a few predators, and mites with mouthparts specialized for filter feeding in water. Some feed on algae in the intertidal zone, in water-filled tree holes, or in sap exudates. Several groups inhabit the nests of insects, mammals, and birds as ectoparasites or commensals. The specialized deutonymphs in several families are parasitic in the hair follicles of mammals. The insect hosts of some species produce acarinaria. One group lives as commensals on the gills of hermit crabs. A large radiation of astigmatic mites lives on feathers and within quills of birds as well as in the avian respiratory tract and air sacs. Another group has radiated among mammals as parasites of the skin, hair, follicles, respiratory passages, ears, and even (rarely) the digestive tract. Some astigmatic mites feed in dung or carrion, and others are important pests of stored food products, including not only grains but also stored meat and dried fish. This order also includes house dust mites (*Dermatophagoides* spp.), the itch or scabies mite (*Sarcoptes scabiei*), and mange mites.

Acknowledgments

We gratefully thank Valerie Behan-Pelletier, Clifford Desch, Mark Harvey, Gustavo Hormiga, Adriano Kury, Evert Lindquist, Roy Norton, Barry OConnor, Lorenzo Prendini, Heather Proctor, Jeffrey Shultz, William Shear, and David Walter for comments on the manuscript, whole or in part. Estimates of mite species numbers are based on Walter and Proctor (1999).

See Also the Following Articles

ARTHROPODS, AMAZONIAN • CRUSTACEANS • INSECTS, OVERVIEW • INVERTEBRATES, FRESHWATER, OVERVIEW • INVERTEBRATES, MARINE, OVERVIEW • INVERTEBRATES, TERRESTRIAL, OVERVIEW • MYRIAPODS

Bibliography

Adis, J., and Harvey, M. S. (2000). How many Arachnida and Myriapoda are there worldwide and in Amazonia? *Stud. Neotropical Fauna Environ.*, in press.

Bruin, J., van der Geest, L. P. S., and Sabelis, M. W. (Eds.) (1999). *Ecology and Evolution of the Acari.* Kluwer, New York.

Coddington, J. A., and Levi, H. W. (1991). Systematics and evolution of spiders (Araneae). *Annu. Rev. Ecol. Syst.* 22, 565–592.

Evans, G. O. (1992). *Principles of Acarology.* CAB. International, Wallingford, UK.

Foelix, R. F. (1996). *Biology of Spiders,* 2nd ed. Oxford Univ. Press, Oxford.

Harvey, M. S. (1992). The phylogeny and systematics of the Pseudoscorpionida (Chelicerata: Arachnida). *Invertebrate Taxonomy* 6, 1373–1435.

Houck, M. A. (1994). *Mites: Ecological and Evolutionary Analyses of Life History Patterns.* Chapman & Hall, New York.

Johnston, D. E., Kethley, J. B., and OConnor, B. M. (1982). Acari. In *Synopsis and Classification of Living Organisms* (S. P. Parker, Ed.), pp. 111–169. McGraw-Hill, New York.

Kaestner, A., Levi, H. W., and Levi, L. R. (1980). *Invertebrate Zoology,* Vol. 2. Krieger, New York.

Krantz, G. W. (1978). *A Manual of Acarology.* Oregon Univ. Book Stores, Corvallis.

Parker, S. P. (Ed.) (1982). *Synopsis and Classification of Living Organisms,* Vol. 2. McGraw-Hill, New York.

Polis, G. (Ed.) (1990). *The Biology of Scorpions.* Stanford Univ. Press, Stanford, CA.

Punzo, F. (1998). *The Biology of Camel Spiders (Arachnida: Solifugae).* Kluwer, New York.

Shear, W. A. (1999). *Introduction to Arthropoda and Cheliceriformes. Microscopic Anatomy of Invertebrates.* Vol. 8A: *Chelicerate Arthropoda,* pp. 1–19. Wiley-Liss, New York.

Shultz, J. W. (1990). Evolutionary morphology and phylogeny of Arachnida. *Cladistics* 6, 1–38.

Stockwell, S. A. (1989). Revision of the phylogeny and higher classification of scorpions (Chelicerata). PhD dissertation, University of California, Berkeley.

Treat, A. E. (1975). *Mites of Moths and Butterflies.* Comstock, Ithaca, NY.

Walter, D. E., and Proctor, H. C. (1999). *Mites: Ecology, Evolution and Behaviour.* Univ. of New South Wales Press and CABI, Wallingford.

Weygoldt, P. (1996). The evolutionary morphology of whip spiders: Towards a phylogenetic system (Chelicerata: Arachnida: Amblypygi). *J. Zool. Syst. Evol. Res.* 34, 185–202.

Wheeler, W. C., and Hayashi, C. Y. (1998). The phylogeny of the extant chelicerate orders. *Cladistics* 14, 173–192.

Wrensch, D. L., and Ebbert, M. A. (Eds.) (1993). *Evolution and Diversity of Sex Ratios in Insects and Mites.* Chapman & Hall, New York.

ARCHAEA, ORIGIN OF

Costantino Vetriani
Rutgers University

GLOSSARY

acidophiles From the Latin *acidus* (sour) and the Greek *philos* (loving). Includes organisms that grow optimally at low pH.

Archaea One of three domains of life. From the Greek *archaios* (ancient, primitive). Prokaryotic cells; formerly called Archaeobacteria.

Bacteria One of three domains of life. From the Greek *bacterion* (staff, rod). Prokaryotic cells; formerly called Eubacteria.

Crenarchaeota One of two kingdoms of organisms of the domain Archaea. From the Greek *crene* (spring, fountain), for the resemblance of these organisms to the ancestor of the Archaea, and *archaios* (ancient). Include sulfur-metabolizing, extreme thermophiles.

Eukarya One of three domains of life. From the Greek *eu* (good, true) and *karion* (nut; refers to the nucleus). Eukaryotic cells; formerly called Eucaryotes.

Euryarchaeota One of two kingdoms within the domain Archaea. From the Greek *eurys* (broad, wide), for the relatively broad patterns of metabolism of these organisms, and *archaios* (ancient). Includes halophiles, methanogens, and some anaerobic, sulfur-metabolizing, extreme thermophiles.

halophiles From the Greek *halos* (salt) and *philos* (loving). Includes organisms that grow optimally at high salt concentrations.

hyperthermophiles From the Greek *hyper* (over), *therme* (heat), and *philos* (loving). Includes organisms that grow optimally at temperatures higher than 80°C.

Korarchaeota Proposed third kingdom within the domain Archaea. From the Greek *koros* (young man), for the early divergence of this group during the evolution of the Archaea, and *archaios* (ancient). Includes a small group of ribosomal RNA sequences retrieved from geothermally heated sediments.

mesophiles From the Greek *mesos* (middle) and *philos* (loving). Includes organisms that grow optimally at temperatures between 20 and 50°C.

methanogens Strictly anaerobic Archaea that produce (Greek *gen*: to produce) methane.

phylogeny The study of the evolutionary relationships among organisms or genes.

ribosomal rRNA Universally distributed molecule among cellular life forms. Widely used to infer the evolutionary relationships among organisms.

thermophiles From the Greek *therme* (heat) and *philos* (loving). Includes organisms that grow optimally at temperatures between 50 and 80°C.

ARCHAEA (FORMERLY ARCHAEOBACTERIA) ARE A GROUP of microscopic organisms that constitute a phy-

Encyclopedia of Biodiversity, Volume 1

logenetically separate doman of life distinct from the other two domains, the Bacteria (Eubacteria) and Eucarya (Eukaryotes). Archaea are generally, although not exclusively, found living in extreme terrestrial or aquatic environments such as hot springs, deep-sea thermal vents, or under conditions of extreme pH and/or concentrated salinity. Because of their adaptation to unusual environments, modern Archaea may provide insights into the early history of life on Earth.

I. INTRODUCTION

A. Historical Background

For more than 50 years, the notion that there were two basic kinds of living organisms, prokaryotes and eukaryotes, was generally accepted. The placement of a specific organism within one or the other group was based on overall morphological and phenotypic similarities. In 1965, Zuckerkandl and Pauling for the first time suggested that sequences of molecules could be used to reconstruct evolutionary history. Thus, they opened the way to molecular phylogeny, a discipline that studies the evolutionary relationships among organisms or genes. Using molecular phylogeny, in the late 1970s Carl Woese and co-workers found evidence that life consisted not of two but of three distinct groups of organisms: eukaryotes and two kinds of prokaryotes, the Eubacteria and the Archaeobacteria. Their evidence was based on the phylogenetic analysis of a single molecule, the small-subunit ribosomal RNA (SSU rRNA), which is now generally considered an excellent molecule for studying the evolutionary relationships among organisms. The tripartite view of life was formally proposed by Woese *et al.* in 1990 in the form of three domains: the Eucarya (formerly Eukaryotes), Bacteria (formerly Eubacteria), and Archaea (formerly Archaeobacteria).

Although Archaea resemble Bacteria morphologically, they differ in a variety of cellular and genetic features. From a cellular and biochemical standpoint, Archaea have some unique characteristics, such as the composition of their cell wall and membranes. From a genetic standpoint, Archaea have a unique combination of characteristics once thought to be exclusive to either the Bacteria or Eucarya. Recent analysis of genomic sequences revealed that there is a tendency for archaeal genomes to be "grab-bags" or chimeras of both bacterial and eukaryotic sequences.

Since many Archaea thrive under unusual environmental conditions that are lethal to most organisms, such as high temperature, high salinity, or extreme pH, they provide experimental models to study adaptations to extreme environments. From an evolutionary standpoint, the study of modern Archaea may offer valuable insights into the nature of the evolution of biological processes and the origin of life.

B. Taxonomy and Phylogeny

Molecular phylogenetic studies involving the comparison of SSU rRNA sequences revealed that the domain Archaea consists of two kingdoms: the Crenarchaeota and the Euryarchaeota (Fig. 1). Physiologically characterized members of the Crenarchaeota include extremely thermophilic microorganisms belonging to the order Thermoproteales (e.g., *Thermoproteus* and *Pyrobaculum*) and to the proposed order Igneococcales (e.g., *Pyrodictium* and *Desulfurococcus*), respectively, and thermoacidophilic microorganisms belonging to the order Sulfolobales (e.g., *Sulfolobus* and *Acidianus*). The Euryarchaeota comprise a metabolically versatile group, which includes all the methanogenic Archaea (orders Methanopyrales, Methanococcales, Methanobacteriales, Methanomicrobiales, and Methanosarcinales), the extreme halophiles (order Halobacteriales), the extreme thermophiles (orders Thermococcales and Archaeoglobales), and the extremely acidophilic microorganisms belonging to the order Thermoplasmales.

The recent application of molecular approaches to survey microbial diversity in natural environments has provided a new perspective on the distribution and evolutionary relationships of the Archaea. A small group of SSU rRNA sequences retrieved from geothermally heated sediments were found to be different enough from all other archaeal rRNA sequences to warrant the proposal of a new kingdom within the Archaea, namely, the Korarchaeota (see Fig. 3). Furthermore, the Crenarchaeota, once thought to be restricted to high-temperature environments, have been found to be ubiquitously distributed, challenging earlier evidence that these microorganisms were confined to high-temperature ecological niches.

II. DIVERSITY, ECOLOGY, AND PHYSIOLOGY OF THE ARCHAEA

In general, the Archaea fall within three general physiological types: (i) thermophiles that live at high tempera-

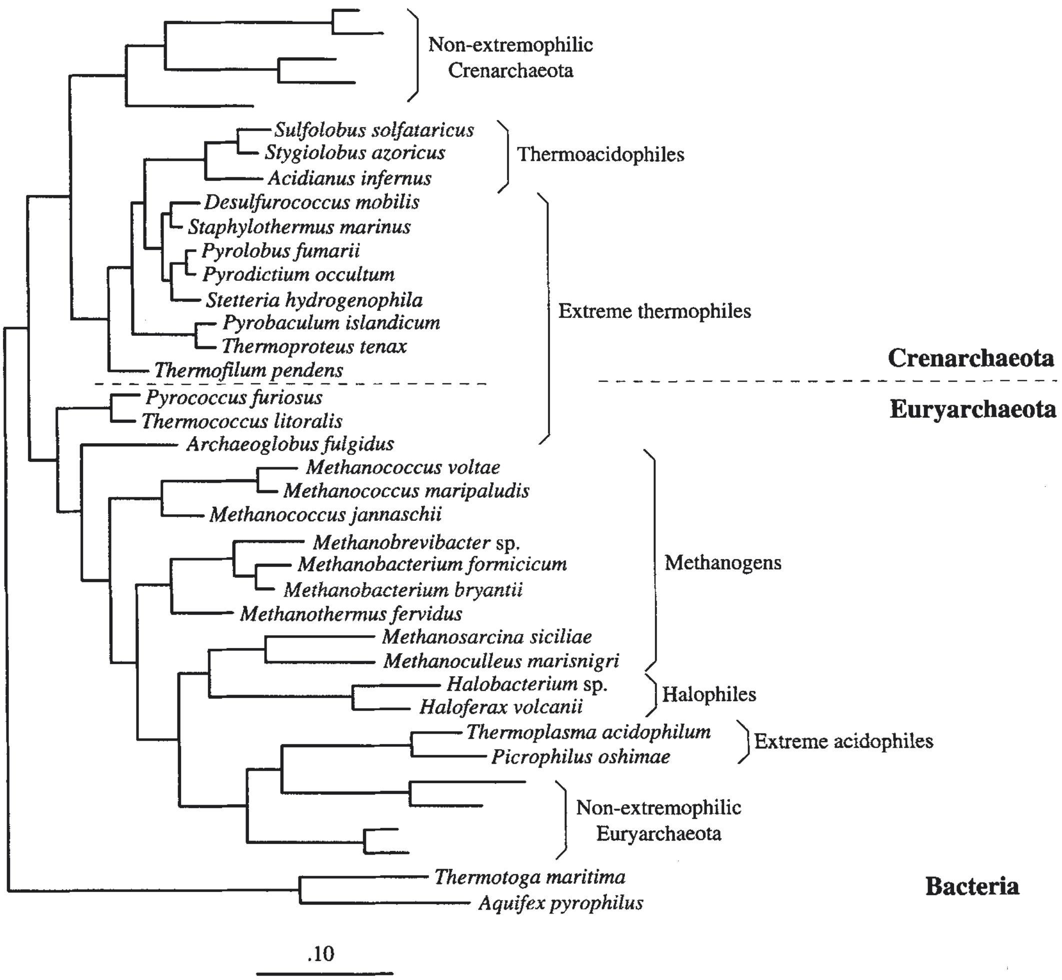


FIGURE 1 Unrooted phylogenetic tree showing the relationships among representative species of Archaea. Branch lengths (indicated by horizontal distances from nodes) are proportional to the evolutionary distances determined by comparative analyses of SSU rRNA sequences. The scale represents the expected number of changes per sequence position.

tures (some thermophiles are adapted to live in extremely low pH conditions and are known as thermoacidophiles); (ii) methanogens that inhabit strictly anaerobic environments and convert carbon dioxide and simple organic molecules to methane; and (iii) halophiles that inhabit highly saline environments and are characterized by a chemoorganotrophic, aerobic metabolism. These three major groups of Archaea are adapted to live in unusual environments and are commonly known as extremophiles. However, molecular surveys have recently revealed that Archaea are much more diverse than previously known, and that they are also widespread in more common biotopes. Nonextremophilic Archaea have been found in marine plankton, terrestrial soils, and freshwater and marine sediments. Because these newly discovered Archaea have resisted cultivation to date, their physiology and metabolic potential are unknown.

A. Thermophilic Archaea

Extremely thermophilic Archaea, or hyperthermophiles, comprise a group of microorganisms which are adapted to grow at temperatures higher than 80°C. Hyperthermophilic Archaea are generally restricted to environments in which geothermal energy is available, such as hot springs, solfataras, geothermally heated marine sediments, and submarine hydrothermal vents.

These environments are rich in sulfur and sulfides, and consequently many of the thermophiles have a sulfur-dependent metabolism. Extremely thermophilic Archaea can carry out a variety of respiratory processes and in most cases elemental sulfur is used as either an electron donor or an electron acceptor. Elemental sulfur is formed from geothermal H_2S either by spontaneous oxidation of H_2S with O_2 or through the reaction between H_2S and SO_2. Submarine volcanic environments include shallow (2–50 m) and deep-sea hydrothermal vents (to depths exceeding 3700 m), where the pressure, even at shallow depths, can raise the boiling point of water sufficiently to select for organisms capable of growth at temperatures higher than 100°C.

Among the Crenarchaeota, *Pyrodictium* and *Pyrolobus* (order Igneococcales) are chemolithotrophic sulfur-dependent hyperthermophiles whose maximum growth temperatures of 110 and 113°C, respectively, represent the upper temperature limits for life known so far. *Pyrodictium* is a strict anaerobe and grows on H_2 and S^0. *Pyrolobus* is unusual in that it is capable of reducing both NO_3^- and $S_2O_3^{2-}$ to NH_4^+ and H_2S, respectively, with H_2 as the electron donor. *Desulfurococcus* and *Staphylothermus* (order Igneococcales) are phylogenetically clearly separate from the *Pyrodictium* group (Fig. 1). These coccoid or disc-shaped organisms have an optimal growth temperature higher than 85°C and, in contrast with the *Pyrodictium* group, a maximum temperature not higher than 100°C. They can grow chemolithoautotrophically by sulfur reduction to H_2S or heterotrophically by sulfur respiration of various organic substrates. *Thermoproteus* and *Thermofilum* (order Thermoproteales) are rod-shaped hyperthermophiles that grow in mildly acidic conditions at temperatures up to 95°C. They are both strict anaerobes that can grow chemolithotrophically on H_2 or chemoorganotrophically on complex carbon substrates with S^0 as an electron acceptor. *Pyrobaculum aerophilum* (order Thermoproteales) is a rod-shaped hyperthermophile capable of aerobic respiration in the presence of very low oxygen concentrations (~0.3%) and nitrate reduction under strictly anaerobic conditions.

Solfataras are terrestrial volcanic environments which abound in sulfur and are gassed by emanating steam carrying CO_2, H_2, and H_2S, with temperatures up to 100°C. Thermoacidophilic members of the order Sulfolobales inhabit acidic solfataras and oxidize H_2S or S^0 to H_2SO_4 using organic carbon or fixing CO_2 as a carbon source. *Sulfolobus* grows at, or close to, the surface of acidic solfataras at temperatures between 75 and 85°C and low pH. No thermoacidophilic, aerobic sulfur oxidizers of the *Sulfolobus* type have been isolated from deep-sea hydrothermal vents. This is related to the steep gradient that occurs between the rising hot, anoxic hydrothermal fluid and the cold, oxygenated ambient seawater, which does not provide an ideal niche for these type of organisms. *Acidianus* (order Sulfolobales) is a close relative of *Sulfolobus* that grows at an optimum temperature of 85°C and pH 2. *Acidianus* is a more versatile organism than *Sulfolobus* in that it is capable of growth under both aerobic and anaerobic conditions. During aerobic growth, S^0 is oxidized to H_2SO_4, whereas under anaerobic conditions H_2 is oxidized and S^0 is concomitantly reduced to H_2S.

Among the Euryarchaeota, *Pyrococcus* and *Thermococcus* (order Thermococcales) are two closely related, coccoid, sulfur-reducing hyperthermophiles that differ primarily in their optimal growth temperatures (100 and 88°C, respectively) (Fig. 2). These organisms are obligate anaerobic chemoorganotrophs that utilize proteins and other complex organic mixtures and reduce S^0 to H_2S. *Archaeoglobus* (order Archaeoglobales) is an extremely thermophilic, sulfate-respiring organism which has been isolated from shallow and deep-sea hydrothermal environments. *Archaeoglobus* grows at temperatures up to 95°C using H_2, lactate, or complex organic mixtures as electron donors and reducing SO_4^{2-} and $S_2O_3^{2-}$ to H_2S. *Ferroglobus* (order Archaeoglobales) is a coccoid, hyperthermophilic organism capable of oxidizing Fe^{2+} at neutral pH under anoxic conditions.

The order Thermoplasmales includes two moderately thermophilic organisms (optimum temperature growth approximately 60°C), *Thermoplasma* and *Picrophilus*, isolated from coal refuse piles and acidic solfataras. *Thermoplasma* is an acidophilic, cell wall-less Archaeon that resembles bacterial *Mycoplasma* species. It can grow aerobically and anaerobically. Under anaerobic conditions there is a requirement for S^0, which is reduced to H_2S. *Picrophilus* is an extremely acidophilic, heterotrophic organism adapted to grow optimally at pH 0.7. Table I summarizes some characteristics of the thermophilic Archaea.

B. Methanogenic Archaea

Methanogenic *Archaea* are characterized by their metabolic capability of producing methane. Methanogenesis is a strictly anaerobic respiratory means of metabolism that produces cellular energy in the form of ATP through the reduction of carbon dioxide, formate, or CO (CO_2-reducing methanogens), methanol or methylamines (methylotrophic methanogens), or acetate

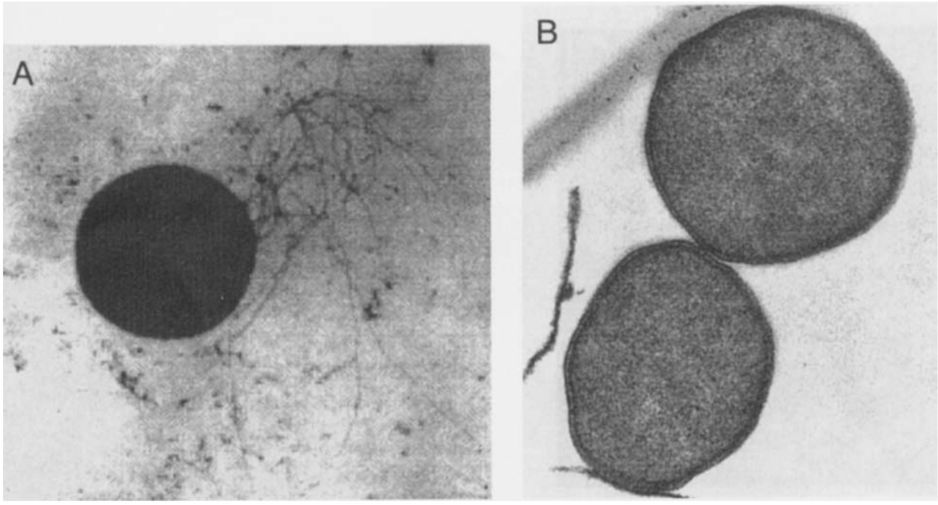


FIGURE 2 Electron micrographs of extremely thermophilic Archaea. (A) Platinum-shadowed *Thermococcus peptonophilus* cell showing the flagella (reproduced with permission from Gonzalez *et al.*, 1995). (B) Pyrococcus horikoshii cells (courtesy of J. M. Gonzalez).

(aceticlastic methanogens), respectively, to methane. Typical methanogenic reactions are

$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O$	(CO_2-reducing methanogens)
$4CH_3OH \rightarrow 3CH_4 + CO_2 + 2H_2O$	(Methylotrophic methanogens)
$CH_3COOH \rightarrow CH_4 + CO_2$	(Acetoclastic methanogens)

Most methanogens are capable of autotrophic growth using molecular hydrogen to reduce CO_2 in a multistep reaction that requires coenzymes unique to this group of organisms, such as methanofuran, methanopterin, F_{420}, F_{430}, and coenzyme M. Methanofuran is a low-molecular-weight phenol derivative involved in the first step of methanogenesis. Methanopterin is a pterin-containing coenzyme which resembles the vitamin folic acid and serves as a C1 carrier during the reduction of CO_2 to CH_4. Coenzyme F_{420} is a flavin derivative involved in redox reactions. The oxidized form of F_{420} has a characteristic blue-green fluorescence at 420 nm and is very useful in recognizing methanogens microscopically. Coenzymes M and F_{430} play a crucial role in the final step of the methanogenesis as part of the methyl reductase system. F_{430} is a yellow, soluble, nickel-containing tetrapyrrole that, unlike F_{420}, does not fluoresce.

All methanogenic Archaea are strictly anaerobic and occupy anoxic habitats, such as sediments, marshes, waterlogged soils, submarine hydrothermal vents, and the digestive tracts of animals. Methanogens are usually abundant in environments depleted in Fe^{3+}, NO^{3-}, and SO_4^{2-} because these substrates stimulate growth of anaerobic bacteria that can outcompete them. Therefore, methanogenesis is usually limited in sulfate-rich marine environments because the methanogens have to compete with sulfate-reducing bacteria for H_2. Since oxidation of H_2 with sulfate as the electron acceptor is thermodynamically more favorable than when CO_2 is the electron acceptor (as in methanogenesis), sulfate reducers are usually favored. However, since methylotrophic methanogens can use noncompetitive substrates (such as methylamines, which are inaccessible to sulfate-reducers), they can be detected in some sulfate-rich environments. Most methanogens have been found to contain the nitrogenase protein complex; therefore, they are also capable of anaerobic nitrogen fixation under certain growth conditions.

The methanogenic Archaea is a diverse group with many species; it is the largest taxonomic group of Archaea. The order Methanobacteriales includes mainly rod-shaped methanogens that grow by CO_2 reduction, although a few species of the genus *Methanosphaera* are cocci that grow only by using H_2 to reduce methanol to methane. *Methanobrevibacter* species are very short rods which use H_2 or formate to reduce CO_2 to CH_4. The genus *Methanothermus* includes extremely thermophilic methanogens isolated from terrestrial geothermal springs. They can grow at 88°C on CO_2 and H_2.

TABLE I

Habitat and General Characteristics of Thermophilic *Archaea*

Kingdom Order *Genus*	Habitat	Morphology	Temperature (°C)		Optimum pH
			Optimum	Maximum	
Crenarchaeota					
Igneococcales					
Pyrodictium	Coastal and deep-sea hydrothermal vents	Disc-shaped	105	110	6
Pyrolobus	Deep-sea hydrothermal vent, Mid-Atlantic Ridge	Lobed cocci	105	113	5.5
Desulfurococcus	Hot springs and solfataras	Sphere	85	95	6
Staphylothermus	Submarine hydrothermal vent, Italy	Spheres in clumps	92	98	6–7
Thermoproteales					
Thermoproteus	Hot springs and solfataras	Rod	88	96	6
Thermofilum	Solfataras	Rod	88	95	5.5
Pyrobaculum	Submarine hydrothermal vents and solfataras	Rod	95	100	6
Sulfolobales					
Sulfolobus	Acidic solfataras	Lobed sphere	75–85	87	1–5
Acidianus	Acidic solfataras	Sphere	85–90	95	2
Desulfurolobus	Acidic solfataras	Lobed sphere	80	87	2.5
Stygiolobus	Hot spring, Azores	Lobed sphere	80	89	3
Euryarchaeota					
Thermococcales					
Pyrococcus	Coastal and deep-sea hydrothermal vents	Sphere	100	106	6–8
Thermococcus	Coastal and deep-sea hydrothermal vents	Sphere	88	98	6–7
Archaeoglobales					
Archaeoglobus	Coastal and deep-sea hydrothermal vents	Cocci	83	95	7
Ferroglobus	Submarine hydrothermal vent, Italy	Cocci	85	95	7
Thermoplasmales					
Thermoplasma	Coal refuse pile, solfataras	Sphere, filaments	59	63	2
Picrophilus	Acidic solfatara, Japan	Cocci	60	65	0.7

The order Methanomicrobiales encompass a genetically diverse group of CO_2-reducing methanogens that grow at temperatures ranging from 30 to 60°C.

The order Methanosarcinales includes morphologically diverse organisms such as *Methanosarcina*, which form irregular spheroid bodies occurring alone or in aggregates of cells, and *Methanothrix*, which forms sheathed rods. Both *Methanothrix* and *Methanosarcina* are acetoclastic methanogens, although the latter can also use methanol and methylamines as substrates for methanogenesis.

The order Methanococcales includes mesophilic and extremely thermophilic methanogens characterized by irregular coccoid morphology. *Methanococcus jannaschii* was isolated from a deep-sea hydrothermal vent system where it grows at 88°C on CO_2 and H_2 released in volcanic gases.

Methanopyrus (order Methanopyrales) is an extremely thermophilic methanogen isolated from hydrothermally influenced sediments, and it is capable of autotrophic growth at temperatures up to 110°C. Table II summarizes some characteristics of the methanogenic Archaea.

C. Halophilic Archaea

The halophilic Archaea occur in environments characterized by high salinity. Salt-neutral lakes, saline soils, solar evaporation marine salterns (areas where sea salt is produced), and subsurface haline deposits are among the most common habitats for halophiles. Most of the halophilic Archaea are red or orange because of the presence of carotenoid pigments in the cell envelope. Frequently, due to their abundance, they impart a red color to the brine. Alkaliphilic halophiles are found in soda lakes, which are highly alkaline environments

TABLE II

General Characteristics of Methanogenic *Archaea*

Order *Genus*	Morphology	Substrates	Temperature (°C)
Methanobacteriales			
Methanobacterium	Long rods	$H_2 + CO_2$, formate	35–40
Methanobrevibacter	Short rods	$H_2 + CO_2$, formate	30–38
Methanosphaera	Cocci	Methanol + H_2 (both needed)	36–40
Methanothermus	Rods	$H_2 + CO_2$	83–88
Methanococcales			
Methanococcus (mesophilic sp.)	Irregular cocci	$H_2 + CO_2$, pyruvate + CO_2, formate	35–40
Methanococcus (thermophilic sp.)	Irregular cocci	$H_2 + CO_2$	88
Methanomicrobiales			
Methanomicrobium	Short rods	$H_2 + CO_2$, formate	40
Methanogenium	Irregular cocci	$H_2 + CO_2$, formate	30–57
Methanospirillum	Spirilla	$H_2 + CO_2$, formate	30–40
Methanoplanus	Plate-shaped cells	$H_2 + CO_2$, formate	32–40
Methanoculleus	Coccus	$H_2 + CO_2$, formate	37–60
Methanosarcinales			
Methanosarcina	Large irregular cocci in packets	$H_2 + CO_2$, methanol, methylamines, acetate	35–50
Methanolobus	Irregular cocci in aggregates	Methanol, methylamines	30–40
Methanohalobium	Irregular cocci	Methanol, methylamines	50
Methanococcoides	Irregular cocci	Methanol, methylamines	23–35
Methanohalophilus	Irregular cocci	Methanol, methylamines, methyl sulfides	26–36
Methanothrix	Long rods to filaments	Acetate	35–60
Methanopyrales			
Methanopyrus	Rods in chains	$H_2 + CO_2$	100

whose high pH (8 to >12) is due to high levels of carbonate. Halophiles require at least 1.5 *M* NaCl for growth, and optimal salt concentrations are usually in the range of 2–4 *M* NaCl.

Neutrophilic and alkaliphilic halophiles both belong to the order Halobacteriales. The members of the neutrophilic group, represented by the genera *Halobacterium, Halococcus, Haloarcula*, and *Haloferax*, grow optimally under conditions of high magnesium and sodium concentrations (0.5 and 4.0 *M*, respectively) at pH ranging from 5 to 8. Alkaliphilc halophiles (genera *Natronobacterium* and *Natronococcus*) grow optimally at low magnesium concentrations and pH of approximately 10.

Halophilic Archaea have evolved several physiological adaptations that permit their growth in habitats with salt concentrations that cause cellular dehydration and protein denaturation in other organisms. Halophilic Archaea such as *Halobacterium* resist high salt concentrations by pumping large amounts of K^+ from the environment into the cell such that the concentration of K^+ inside the cell is greater than the concentration of Na^+ outside the cell. This mechanism allows *Halobacterium* to remain in positive water balance and avoid dehydration.

The halophilic Archaea are aerobic and grow heterotrophically using carbohydrates, alcohols, organic acids, and amino acids. *Halobacterium* species are normally aerobic but can grow anaerobically in the presence of light. Under oxygen-limiting conditions and in the presence of light, *Halobacterium* inserts large amounts of a protein called bacteriorhodopsin into the cytoplasmic membrane. This purple pigment, which adsorbs light strongly at approximately 570 nm, acts as a light-driven proton pump and leads to the establishment of an electrochemical membrane potential. Its equilibration can be accompanied by the generation of ATP. Thus, bacteriorhodopsin facilitates a special kind of photophosphorylation. The energy obtained by this mechanism complements that obtained from aerobic substrate oxidation. Table III summarizes some characteristics of the halophilic Archaea.

TABLE III

Habitat and General Characteristics of Halophilic *Archaea*

Order *Genus* *Species*	Habitat	Morphology	Physiological characteristics
Halobacteriales			
Halobacterium		Rod-shaped	
H. salinarum	Salted fish, hypersaline lakes		Amino acids, phototrophic
Haloferax		Rod/disk-shaped	
H. mediterranei	Salterns		Carbohydrates
H. denitrificans	Salterns		Nitrate respiration
Halorubrum		Rod-shaped	
H. lacusprofundi	Deep lake, Antarctica		Carbohydrates
H. sodomense	Dead Sea		Phototrophic
Haloarcula		Pleiomorphic	
H. vallismortis	Death Valley		Carbohydrates
H. marismortui	Dead Sea		Nitrate respiration
Halococcus		Cocci	
H. morrhuae	Dead Sea		Amino acids, Nitrate respiration
H. saccharolyticus	Salterns		Carbohydrates
Natronobacterium		Rod-shaped	
N. gregoryi	Soda lakes		Organic acids, pH 9.5–10
Natronococcus		Cocci	
N. occultus	Soda lakes		Amino acids, pH 9.5–10
Natrialba		Rod-shaped	
N. asiatica	Sea sand		Nonpigmented, extreme halophile

D. Nonextremophilic Archaea

The application of SSU rRNA-based molecular techniques to survey natural microbial populations has recently revealed that Archaea are virtually ubiquitous. Therefore, our view of the Archaea as highly specialized microorganisms adapted to survive in extreme environments is gradually changing to that of a very diverse group of organisms, including more moderate representatives.

Non-extremophilic Archaea have been identified in a variety of environments, such as temperate and cold marine planktonic habitats, freshwater sediments, soils, terrestrial subsurface environments, and permanently cold deep-sea sediments, and in association with several marine metazoans. To date, none of the non-extremophilic Archaea have been isolated in pure culture; therefore, their phenotypic and physiological characteristics remain unknown. However, their ubiquitous distribution and high phylogenetic diversity suggests a wide range of ecological and physiological adaptations.

The distribution and phylogenetic affiliation of the non-extremophilic Archaea are summarized in Table IV.

III. ORIGIN AND EVOLUTION OF THE ARCHAEA

Archaea occupy a pivotal phylogenetic position between the two other domains of life, Bacteria and Eukarya (Fig. 3). With the recent sequencing of the entire genomes of several Archaea, it has become evident that archaeal genomes share bacterial and eukaryotic features. Furthermore, the availability of an increasing number of gene sequences from organisms belonging to the three domains of life allows the identification of a "universal" set of protein families present in all three domains. This universal set of proteins can then be used as an estimate of the genome content and the cellular processes that were present in the last universal

TABLE IV
Distribution and Phylogenetic Affiliation of Nonextremophilic *Archaea*

Distribution	Phylogenetic affiliation
Marine Habitats	
Surface and deep waters (up to 3000 m)	*Crenarchaeota, Euryarchaeota*
Temperate coastal sediments (12 m)	*Crenarchaeota, Euryarchaeota*
Low-temperature deep-sea sediments (1500 to 4500 m)	*Crenarchaeota, Euryarchaeota*
Temperate microbial mats at deep-sea hydrothermal vent	*Crenarchaeota, Euryarchaeota*
Antarctic low-temperature surface waters	*Crenarchaeota, Euryarchaeota*
Salt marsh	*Euryarchaeota*
Associated with Marine Metazoans	
Gut of abyssal holothurian *Oneirophanta mutabilis* (4870 m)	*Crenarchaeota*
Digestive tract of fish	*Crenarchaeota, Euryarchaeota*
Tissues of sponge *Axinella mexicana* (10–20 m)	Crenarchaeon *Cenarchaeum symbiosum*
Freshwater Habitats	
Lake Sediments	*Crenarchaeota, Euryarchaeota*
Terrestrial Habitats	
Soils	*Crenarchaeota, Euryarchaeota*
Subsurface paleosol (188 m)	*Crenarchaeota*
Contaminated aquifer	*Crenarchaeota, Euryarchaeota*
Rice roots	*Crenarchaeota, Euryarchaeota*

ancestor of living organisms. However, factors such as horizontal gene transfers between distantly related groups and gene loss make it difficult to draw conclusions about the biology of the last universal ancestor, and different theories have been proposed.

The following sections describe different traits of the Archaea which can be used to infer some properties of the last universal ancestor as well as to provide insights of the origin and evolution of the Archaea.

A. Early Respiratory Processes

The presence of genes encoding for homologous proteins in organisms belonging to both domains Archaea and Bacteria allows for the reconstruction of cellular processes that were present early in the history of life on Earth. In particular, terminal oxidases belonging to oxygen, nitrate, sulfate, and sulfur respiratory pathways are present in both domains.

Cytochrome oxidase, which catalyzes the reduction of oxygen to water and acts as a redox-linked proton pump, is the key enzyme of aerobic metabolism. This enzyme is present in members of both the Crenarchaeota (*Sulfolobus, Acidianus*, and *Pyrobaculum*) and the Euryarchaeota (*Halobacterium* and *Natronobacterium*) as well as in many thermophilic Bacteria (e.g., *Thermus* and *Aquifex*). It is plausible that in the early atmosphere, prior to the advent of oxygenic photosynthesis, microaerophilic organisms similar to the present-day archaeon *Pyrobaculum* may have been able to thrive under extremely low oxygen concentrations, known to have existed as a result of the photolysis of water.

The ability to use sulfate as a terminal electron acceptor is a characteristic common to many mesophilic and thermophilic bacterial species and to the thermophilic archaeon *Archaeoglobus*. In addition, a sulfite reductase-type protein has been identified in *Pyrobaculum*. The enzyme dissimilatory sulfite reductase catalyzes the six-electron reduction of (bi)sulfite to sulfide, which is the central energy-conserving step of sulfate respiration. Phylogenetic evidence suggests an ancestral origin of sulfate respiration, a finding which is consistent with the notion that sulfates of magmatic origin were common in the Archaean time. In addition to sulfate reduction, *Archaeoglobus* produces very small quantities of methane. Factor F_{420} and methanopterin, two coenzymes involved in methanogenesis, are also present in *Archaeoglobus*, although this organism appears to lack other cofactors normally present in methanogens. This apparently intermediate type of metabolism between methanogenesis and sulfur reduction suggested that *Archaeoglobus* may represent a transitional form in the diversification of Archaea from sulfur-respiring to methane-producing and halophilic Archaea. The branching position of *Archaeoglobus* in the SSU rRNA tree is consistent with this interpretation (Fig. 1).

Sulfur respiration has also been proposed to be an ancestral mode of energetic metabolism on the base of two considerations: (i) Volcanic-derived S^0 was probably one of the most abundant electron acceptors in the early atmosphere, and (ii) the capacity to reduce S^0 to H_2S is common among anaerobic, hyperthermophilic members of both Archaea and Bacteria.

B. Adaptive Features to High Temperature

Archaea exhibit a strong adaptive capacity to extremely high temperatures. Since microorganisms are isothermal, they have to develop strategies to avoid thermal

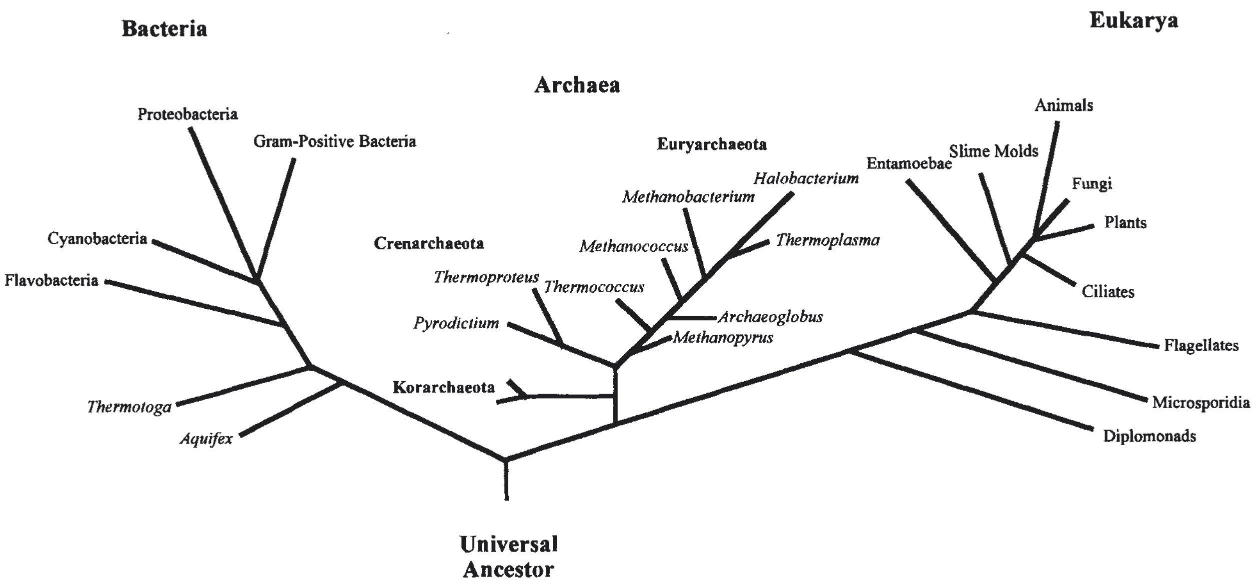


FIGURE 3 Schematic representation of the rooted "universal" SSU rRNA phylogenetic tree showing the position of the Archaea relative to the Bacteria and the Eukarya.

stress. Thus, in order to survive at temperatures in excess of 100°C, hyperthermophiles have to adapt their cell inventory to function optimally at the temperatures in the niches they occupy. Many different mechanisms have been hypothesized by which hyperthermophiles can simultaneously retain stability and plasticity at high temperatures.

One important mechanism of thermoadaptation is the biosynthesis of thermostable proteins and enzymes that maintain sufficient structural integrity to allow optimal catalytic efficiency at high temperatures. Thermostable proteins are often characterized by a very efficient packing density in the hydrophobic core of the molecule as well as by shorter connecting elements (loops) between regions of secondary structure. An elevated number of ionic interactions has frequently been found in highly thermostable proteins. Ionic interactions among monomers are thought to stabilize multimeric proteins at high temperatures. Overall, these characteristics lead to decreased flexibility in the polypeptide chain, a required feature that compensates for increased thermal fluctuations at high temperatures.

Most living beings need the strand-opening potential of negative DNA supercoiling to allow transcription and other DNA-dependent processes. In bacterial nucleoproteins and in eukaryal chromosomes, DNA has a negative superhelicity as a result of topoisomerase activity and wrapping of DNA around histone cores, respectively. Histone-like proteins may play an important role in certain thermophilic Archaea, protecting DNA against thermal denaturation and degradation. The DNA-binding proteins isolated from both *Thermoplasma acidophilum* and *Methanothermus fervidus* stabilize double-stranded DNA molecules *in vitro* by increasing their melting temperatures. A novel class of topoisomerases has been identified in both hyperthermophilic Archaea (e.g., *Sulfolobus*) and Bacteria (e.g., *Thermotoga*). Since these enzymes catalyze positive supercoiling into the DNA, they have been referred to as reverse gyrases. Reverse-gyrase activity is widely distributed in hyperthermophiles and appears to be, at least at critical regions, a requirement to prevent DNA denaturation at high temperatures. Thus, the reverse girase may be considered as an ancestral trait retained by present-day hyperthermophilic Archaea.

C. Evolution

Hyperthermophilic Archaea have been proposed as analogs for the early life on Earth. If life arose and evolved under high-temperature, reduced, sulfur-rich conditions, then the requirement of thermophilic Archaea for high temperature environments as well as their predominantly anaerobic, sulfur-metabolizing phenotype suggest they are good analogs to test such a theory. Two independent lines of evidence suggest that the phenotype of the extant hyperthermophiles may represent remnants of the ancestral Archaea. First, the hyperthermophilic trait is the only one found in both archaeal kingdoms, the Crenarchaeota and the Euryarchaeota. Second, most hyperthermophilic lineages are more slowly evolving than any other archaeal lineage. This is particularly evident when the rate of evolution of the SSU rRNA genes (indicated as the branch length in the phylogenetic tree) is considered (Fig. 1). Mesophilic methanogens, halophiles, and the recently discovered non-extremophilic Archaea appear to evolve much faster than hyperthermophiles (Fig. 1). Furthermore, mesophilic methanogens, such as members of the Methanococcales, evolve more rapidly than their extremely thermophilic relatives. Therefore, the assumption that more slowly evolving lineages (thermophilic) tend to retain more of the ancestral traits than do rapidly evolving lineages (non-thermophilic) suggests a high-temperature ancestry of the Archaea, which later adapted to lower temperatures.

A different theory supporting a non-thermophilic ancestor to extant life forms has been recently formulated on the base of the strong correlation between the Guanine + Cytosine (G + C) nucleotide content of rRNA genes and the optimal growth temperature of microorganisms. The analysis of rRNA genes from hyperthermophiles suggests that survival in high-temperature conditions requires a high G + C content in rRNA genes. However, the estimated G + C contents of ancestral organisms appeared to be incompatible with life at high temperature and suggested a moderate environmental growth temperature for the universal ancestor. This theory leads to the conclusion that present-day thermophily evolved from mesophilic organisms via adaptation to high temperature.

In principle, the nature of an ancestral organism may be inferred from the distribution of homologous traits among its descendants. However, the reconstruction of early evolutionary histories is complicated by phenomena such as horizontal gene transfer (gene swapping) among organisms. As new microbial genomes have become available, it has become increasingly evident that extensive gene swapping, even among distantly related organisms, may have occurred. This tends to blur the history of the early stages of life on Earth; therefore, the nature of the last universal ancestor remains elusive. At the same time, comparative functional genomics is providing a new perspective of the characteristics of the universal ancestor. Because of their evolutionary history, Archaea are likely to play a critical role in our understanding of the early events in the evolution of cells and the origin of life.

See Also the Following Articles

BACTERIAL BIODIVERSITY • BACTERIAL GENETICS • EUKARYOTES, ORIGIN OF • MICROBIAL BIODIVERSITY • ORIGIN OF LIFE, THEORIES OF • PSYCHROPHILES • THERMOPHILES, ORIGIN OF

Bibliography

Aravalli, R. N., She, Q. X., and Garrett, R. A. (1998). Archaea and the new age of microorganisms. *TREE* **13**, 190–194.

Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. (1996). Perspectives on archaeal diversity, thermophily and monophily from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA* **93**, 9188–9193.

Blöch, E., Burggraf, S., Fiala, G., Lauerer, G., Huber, G., Huber, R., Rachel, R., Segerer, A., Stetter, K. O., and Völkl, P. (1995). Isolation, taxonomy and phylogeny of hyperthermophilic microorganisms. *World J. Microbiol. Biotechnol.* **11**, 9–16.

Brown, J. R., and Doolittle, W. F. (1997). Archaea and the Prokaryote-to-Eukaryote transition. *Microbiol. Mol. Biol. Rev.* **61**, 456–502.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., Fitzgerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., and Venter, J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.

Burggraf, S., Huber, H., and Stetter, K. O. (1997). Reclassification of the crenarchaeal orders and families in accordance with 16S rRNA sequence data. *Int. J. Syst. Bacteriol.* **47**, 657–660.

Castresana, J., and Moreira, D. (1999). Respiratory chains in the last common ancestor of living organisms. *J. Mol. Evol.* **49**, 453–460.

Ciaramella, M., Cannio, R., Moracci, M., Pisani, F. M., and Rossi, M. (1995). Molecular biology of extremophiles. *World J. Microbiol. Biotechnol.* **11**, 71–84.

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. USA* **89**, 5685–5689.

Fuhrman, J. A., McCallum, K., and Davis, A. A. (1992). Novel major archaebacterial group from marine plankton. *Nature (London)* **356**, 148–149.

Galtier, N., Tourasse, T., and Gouy, M. (1999). A nonhyperthermophilic common ancestor to extant life forms. *Science* **283**, 220–221.

Gonzalez, J. M., Kato, C., and Horikoshi, K. (1995). *Thermococcus peptonophilus* sp. npv., a fast-growing, extremely thermophilic archaebacterium isolated from deep-sea hydrothermal vents. *Arch. Microbiol.* **164**, 159–164.

Konig, H., and Stetter, K. O. (1989). Archaeobacteria. In *Bergey's Manual of Determinative Bacteriology* (J. T. Staley, M. P. Bryant,

N. Pfennig, and J. G. Holt, Eds.), Vol. 3, pp. 2171–2173. Williams & Wilkins, Baltimore.

Lopez-Garcia, P. (1999). DNA supercoiling and temperature adaptation: A clue to early diversification of life? *J. Mol. Evol.* 49, 439–452.

Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S.-P., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H., and Robb, F. T. (1998). Protein thermostability above 100°C: A key role for ionic interactions. *Proc. Natl. Acad. Sci. USA* 95, 12300–12305.

Vetriani, C., Jannasch, H. W., MacGregor, B. J., Stahl, D. A., and Reysenbach, A.-L. (1999). Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. *Appl. Environ. Microbiol.* 65, 4375–4384.

Wagner, M., Roger, A. J., Flax, J. L., Brusseau, G. A., and Stahl, D. A. (1998). Phylogeny of dissimilatory sulfite reductases supports an early origin of sulfate respiration. *J. Bacteriol.* 180, 2975–2982.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. USA* 87, 4576–4579.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theor. Biol.* 8, 357–366.

ARCTIC ECOSYSTEMS

Terry V. Callaghan,[*,†] Nadya Matveyeva,[‡] Yuri Chernov,[§] and Rob Brooker[¶]
[*]University of Sheffield, [†]Abisko Scientific Research Station, [‡]Komarov Botanical Institute, [§]Severtzov Institute of Evolution, and [¶]Institute of Terrestrial Ecology

GLOSSARY

Arctic Geographically, the region lying to the north of latitude 66.7°N, but environmentally, and in the context of this article, the region to the north of the climatically controlled northern latitudinal treeline which corresponds approximately to the mean July isotherm of 10°C.

permafrost The phenomenon of water which is permanently at or below 0°C. Usually, but not always, the water is in the solid state. The Arctic is characterized by the presence of large, continuous areas of permafrost that have the form of lower soil layers that are permanently frozen with a shallow (usually <1 m) "active layer" which freezes and thaws each year and accommodates belowground biological activity.

refugia In the context of the Arctic, these are land areas which were not covered by ice sheets or glaciers during the last glaciation. Consequently, some biota could survive there and recolonize adjacent areas when the ice retreated. Therefore, these areas are associated with high biodiversity and endemism. Refugia usually occurred in coastal areas (now continental shelves) and ice-free mountain tops or "nunataks."

subarctic The ecotone (ecological boundary zone) connecting the treeless tundras in the north with the taiga or coniferous boreal forests in the south. The area is characterized by the presence of scattered, deciduous or coniferous trees of low stature and is sometimes termed "forest tundra." (Other Arctic vegetation zones are described in the text.)

treeline Here, denotes the northern latitudinal and altitudinal distribution limit of vegetation in which trees are a conspicuous but not necessarily a dominant element. Individual, isolated trees or "oases" of trees in environmentally benign areas are considered to be beyond the treeline.

tundra This is a type of vegetation characteristically occupying the Arctic. However, the term is used in many ways, from characterizing individual plant communities of the Arctic which consist of dwarf shrubs and sedges to characterizing all vegetation above the altitudinal treeline and between the latitudinal treelines and the poles in both hemispheres. In this article, the term is used in the Russian sense to characterize Arctic vegetation lying between the taiga and the region of the polar deserts.

THE ARCTIC REGION is one of the world's last great wilderness areas where human impacts on terrestrial ecosystems are relatively light. Although species diversity is generally lower than at more southerly latitudes, the diversity of animals and plants, communities, and landforms are surprisingly rich. Patterns of biodiversity are strongly coupled with the wide variety of Arctic environments because of relatively low interference by man. The Arctic is therefore far from uniform. This article introduces the reader to current patterns of biodiversity in terrestrial Arctic ecosystems and discusses aspects of the physical, historical, and also biotic environments that have shaped these patterns. We then seek to highlight the importance of the Arctic's biodiversity for ecosystem function and provision of resources for human welfare. Finally, we outline the numerous threats to the Arctic's biota and the major challenges to our further understanding of its responses to change.

I. INTRODUCTION

Although Arctic environments have undergone dramatic changes over millions of years, the twentieth century has been associated with particularly rapid changes in many aspects of the physical environment and particularly with rapid cultural and sociological changes in the north leading to the increasing exploitation and fragmentation of wilderness areas. Because future environmental changes are predicted to be even greater during this century, it is becoming increasingly necessary to document, monitor, and understand the biological resources of the Arctic. Understanding the patterns, causes, and consequences of biodiversity as highlighted by Chapin and Körner (1995) throughout the Arctic's lands is an important aspect of this challenge. However, the Arctic should not be seen as a region remote from the more populated regions of the world at lower latitudes: Several hundred million birds of many species overwintering in temperate regions migrate to summer nesting grounds in the Arctic, and the functions of many Arctic ecosystems in sequestering the greenhouse gas carbon dioxide in soils has contributed to cooling the earth's surface since the end of the last ice age.

This article introduces the reader to current patterns of biodiversity in terrestrial Arctic ecosystems and presents further details of the physical, historical, and biotic environments that have shaped these patterns. We then show the importance of the Arctic's biodiversity in terms of ecosystem function and its provision of resources for human life support and welfare. Finally, we show how the Arctic's biota are threatened by numerous factors and discuss the major challenges to further understanding the biota.

II. DEFINITIONS AND GENERAL ENVIRONMENTAL CHARACTERISTICS OF THE ARCTIC

The Arctic is difficult to define, basically because of two confounding issues—the earth's inclination to the sun and environmental, including climatic, conditions (Nutall and Callaghan, 2000). Strictly, the Arctic is that area north of latitude 66.7°North where the sun does not set below the horizon at midnight on midsummer's night and does not rise above the horizon at midday on midwinter's day. This astronomical feature sets a definitive photoperiod for the Arctic and reduces the amount of solar heat absorbed by the earth's surface. Hence, the Arctic is characterized by having long, cold, and dark winters and short, cool summers with the midnight sun visible.

The energy balance of the earth's surface in the north is complicated, however, by the connectivity of the Arctic's oceans and climate with those further south. The transport of heat from southern latitudes into the Arctic in warm ocean currents creates environments which are much warmer than those that can be considered as typically arctic, for example, those in northern Fennoscandia (Fig. 1) which are warmed by the Gulf Stream. In contrast, where returning cold water flows southwards, unusually cold climates exist for particular latitudes: Tundra environments exist in eastern Canada (Fig. 1) at latitudes at which British agriculture and forestry thrive. Proximity to the coast affects terrestrial temperatures by the formation of mists. On land, topography also affects temperatures locally. As one moves closer to the North Pole, progressively smaller differences in topography become more important in determining temperature.

Often, therefore, the definition of the Arctic depends on the subject and the scientific discipline. Because all biological processes depend on chemicophysical reactions, temperature is an important determinant of biological activity. Also, because nearly all food webs ultimately depend on primary producers which require light for photosynthesis, the strong seasonal variations in photoperiod are also important for biological activity. For biologists, therefore, definitions that incorporate the sensitivity of flora and fauna to low temperatures and specific photoperiods are particularly appropriate. Thus, the 10° summer isotherm (where the mean monthly temperature is at or below 10°C) and the northern limit of forests (the latitudinal treeline), which both represent the southern boundary of the "low Arctic" in Western terminology and the "tundra region" in

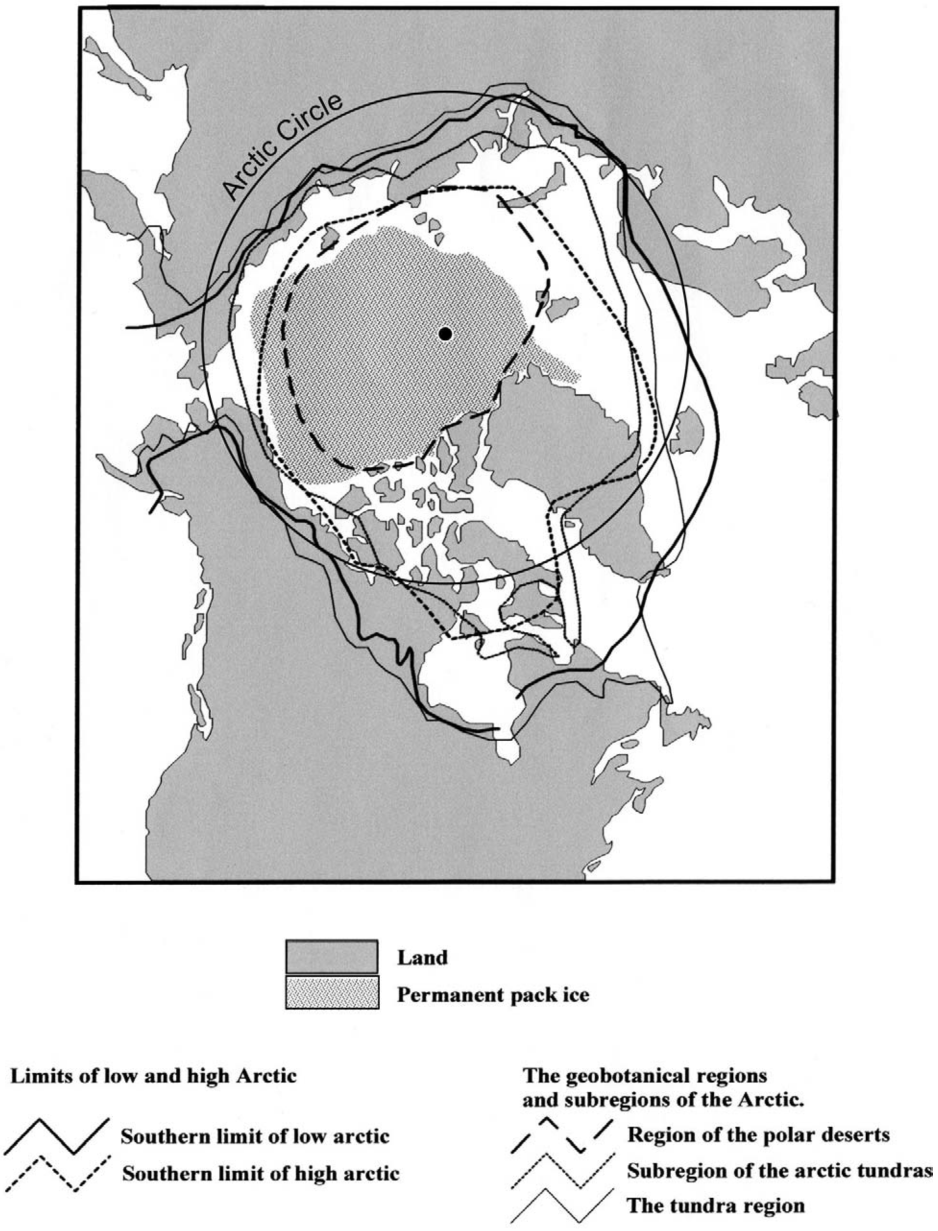


FIGURE 1 Contrasting concepts of vegetation zones of the Arctic according to Bliss and Matveyeva (1992) and Alexsandrova (1980). The low and high Arctic are commonly used in the West, whereas the geobotanical zonation is used in Russia.

Russian terminology, are commonly used to define the Arctic (Fig. 1).

The extent of the Arctic is totally dependent on its definition. Using botanical definitions for the Arctic, Bliss and Matveyeva (1992) calculate that the Arctic comprises approximately 7.5 million km^2, which is approximately 5.5% of the earth's land surface. The Arctic stretches over 33 degrees of latitude from 84°N in Greenland to 51°N in Hudson Bay, Canada (see Jonasson *et al.* in Nutall and Callaghan, 2000). Almost 2 million km^2 is permanent snow and ice. Distances around the Arctic decrease dramatically towards the North Pole. For example, the circumference of the equator is 40,076 km, whereas that of the 70° parallel is only 13,752 km. This has great implications for connectivity between animal and plant populations and species.

The complexity of defining the Arctic is amplified when the Arctic is subdivided into different zones. Although it is clear that the Arctic is not a homogeneous environment, definitions of subzones conflict between Eastern and Western scientific traditions (Bliss and Matveyeva, 1992). In the West, a generally accepted classification system is as follows:

1. Taiga in the south, consisting of predominantly taiga forest, and forest tundra (sparse trees, often with low stature) at its northern edge, near the 10–12°C isotherm for July
2. The low Arctic consisting of tundra vegetation (low thicket-forming shrubs–sedge, tussock–dwarf shrub, and mire communities)
3. The high Arctic consisting of polar semidesert vegetation in the south (cryptogam–herb, cushion plant–cryptogam, and mire communities which do not cover all of the ground) and polar desert (herb-cryptogam communities which cover only approximately 5% of the ground) in the far north where mean July temperatures vary from 6°C in the south to only approximately 2°C in the north. Precipitation in the north is less than 50 mm per year and falls mainly as snow.

In a generally accepted Russian classification, the taiga zone remains the same, but northwards a broad tundra zone classification replaces the low Arctic and polar semidesert subzones of the Western classification. The Russian classification is as follows:

1. The taiga as before
2. The tundra zone, divided into the southern tundra subzone with low shrub–sedge, tussock–dwarf shrub, and mire communities; a typical tundra subzone with sedge–dwarf shrub and ploygonal mire communities; and an Arctic tundra subzone in the north consisting of dwarf shrub–herb communities.
3. The polar desert zone, characterized by cryptogam–herb communities

Despite these classifications, the "subarctic " region is often referred to and represents the ecotone between the tundra and taiga and the forest tundra. In fact, there is a continuous gradient of environmental severity within the Arctic from the boreal forest zone at its southern boundary to the polar deserts of the far north, even if this is interrupted in some places by mountain chains and water bodies.

There are great temperature variations throughout the Arctic. From the southern boundary where the mean July temperature is 10°C and there are more than 1000 degree days >0°C, there is a decrease to the high Arctic where the mean July temperature can be less than 2°C with just over 110 degree days per year. Precipitation also decreases toward the north, from over 1000 mm per year to approximately 50 mm per year. Low temperatures and low precipitation exert critical limitations on the productivity of Arctic ecosystems (Jonasson *et al.* in Nutall and Callaghan, 2000). Net primary production varies from approximately 1 g m^{-2} $year^{-1}$ in polar deserts to 150–800 g m^{-2} $year^{-1}$ in the Subarctic. In the low Arctic, values can range between 100 and 1200. Although there is a good correlation between temperature and productivity, the mechanisms are complex. In many areas, it has been suggested that the most important aspect of temperature is indirect in that it constrains the rate of nutrient cycling in cold arctic soils, thereby resulting in nutrient limitation throughout most of the Arctic (Jonasson *et al.* in Nutall and Callaghan, 2000).

The great spatial differences in temperature regimes throughout the current Arctic are accompanied by even greater variations over time. Arctic landmasses and oceans are relatively young and have been formed by the migrations of landmasses from southerly latitudes. Also, much of the Arctic's land and oceans have been covered even in summer by ice, but only during the past 1.8 million years. The last glacial period, which ended approximately 10,000 years ago in much of the Arctic, displaced terrestrial, freshwater, and marine species southwards from much of the Arctic. A similar process occurred over large lowland areas of the Russian Arctic when they were entirely inundated by the marine transgression (Aleksandrova, 1980; Yurtsev, 1997).

The Arctic's current biota can be seen, therefore, as relatively young and, on land, a depauperate remnant of previous floras and faunas that existed in preglacial eras. For example, forests covered areas of Greenland, Svalbard, and the Canadian high Arctic Archipelago during the Tertiary Period, ending approximately 3 million years ago (this signified a warmer Arctic but also an Arctic landmass that was a few degrees latitude further south). The megafauna, such as the mammoth, of the Russian Arctic and extensive tundra steppe communities disappeared during the early Holocene.

Although much of the current Arctic floras immigrated into the Arctic during the Holocene as the ice retreated, some Arctic areas, "refugia" and "nunataks," remained ice free and supported biota for periods long enough to facilitate the establishment of endemic taxa. The extent to which species colonized the Arctic during the Holocene from such refugia, or from immigration from the south, is unclear. One example of a species thought to have survived in a refugium on Svalbard and then spread out is *Pedicularis dasynatha* (Odasz as quoted by McGraw in Chapin and Körner, 1995).

During the recolonization of the Arctic, the continuity of landmasses allowed a movement of biota into and out of and around the Arctic, e.g., across the Bering land bridge. This connectivity together with the decreasing longitudinal distances toward the pole allowed, and continues to allow, the mobility of many organisms such as marine mammals, migrating caribou, and particularly migratory birds. Such mobility adds considerable complexity to understanding biodiversity in the Arctic because some of the causal factors are operative when migratory organisms are in other latitudes.

Despite this general connectivity, barriers to dispersal of species occur throughout the Arctic. In Alaska, the Brooks mountain range runs approximately east–west and separates the taiga forest zone to the south from the tundra zone in the north. In the Canadian Arctic, much of the Northwest Territories is an archipelago and in Greenland, particularly along the east coast, glaciers flowing to the sea interrupt the continuity of ice-free coastal lands. In Fennoscandia, the mainland is represented mainly by subarctic areas in the mountains, and high Arctic environments are found only on the Svalbard archipelago. In the Russian Arctic, the Ural mountains running north–south separate the vast landmass longitudinally. However, on the Taimyr Peninsula, there is probably the greatest latitudinal continuity of land reaching into the Arctic.

III. PATTERNS OF BIODIVERSITY IN THE ARCTIC

It is not possible to state the exact number of species in the Arctic because of a lack of taxonomic knowledge (particularly for the Protozoa, Nematoda, terrestrial Oligochaeta, many Acari taxa, and insect families in the orders Diptera, Hymenoptera, and Lepidoptera) and incompatible synonyms used in different countries. However, some estimates of general species number can be made (Matveyeva and Chernov in Nutall and Callaghan, 2000; Table I).

The number of species in the majority of plant and animal classes that are of importance in the higher Arctic environments amounts, on average, to approximately 2.5% of the worldwide species number for a given taxon (Table I). The majority of the higher terrestrial faunal taxa are represented in the Arctic, although on a species number basis the Arctic contains less than 1% of the world's fauna. In comparison, the angiosperms, the most advanced plant taxon, have fewer species in the Arctic than the cryptogams (Table I). Thus, the Arctic's biota is not just an extremely depauperate replication of the world's biota but also has a very distinctive structure as exemplified by the proportions of the highest taxa of plants and animals.

TABLE I

Biodiversity Estimates for the Arctic Compared with World Biota[a]

Taxon					
Animals			Plants		
Group	Number	% of world biota	Group	Number	% of world biota
Mammals	55	1.8	Angiosperms	1800	0.7–0.8
Birds	200	3.4	Monocotyledons		0.6
Insects	3000	0.3–0.4	Dycotyledons		0.2
Diptera	1600–1800	1.0	Gymnosperms		1.3–1.7
Beetles	350	0.1	Pteridophytes		0.3
Springtails	400–500	7.0–8.0	Mosses	780	5.7–7.0
Spiders	300	0.9	Liverworts		2.2–3.0
Mites	700	2.0	Lichens	1300	5.0–6.5
Other groups[b]	600	—	Algae	—	—
			Fungi	~5000	—
Total estimate	6000–7000	—			

[a] Modified from Matveyeva and Chernov in Nutall and Callaghan (2000).

[b] Amphibians (4), Centipedes (10), Mollusks (12–24), Worms (70), and Nematodes (~500).

Classes that are well represented in, and are therefore particularly well adapted to, Arctic environments constitute approximately 1–3% or more of global diversity. These include birds, springtails, mosses, and lichens. The largest classes of global biota (e.g., insects and flowering plants) have the lowest level of representation in the Arctic at 0.3–0.4 and 0.8%, respectively. In contrast, relatively small groups such as springtails (*Collembolla*) and horsetails (*Equisetales*) include 7, 8, and 25%, respectively. At lower taxonomic levels—orders, families, and genera—some taxa are endemic to the Arctic. For instance, all species of the order of the diving birds Gaviiformes are found in the Arctic.

Although the proportion of various taxa that are found in the Arctic may be very small, there is a common misunderstanding that this equates to low species numbers in Arctic communities (Chapin and Körner, 1995). This is not necessarily the case. For example, the wide ecological distribution of many Arctic plants, together with their small stature, results in high numbers of species per unit area in certain communities. Thus, on Taymyr, Matveyeva (Matveyeva and Chernov as cited in Nutall and Callaghan, 2000) recorded from 110 to 182 species per sample plot (100 m^2). There were 40–50 species within 1 m^2 and up to 25 species within 1 dm^2. Even in the polar desert, approximately 50–60 species were recorded per 25 m^2 in both the Eurasian and Canadian Arctic.

As with diversity at the level of community composition, low absolute species numbers do not necessarily equate to low functional diversity within communities. Genetic heterogeneity of ecotypes is a reservoir of genotypic variation on which selection acts, fitness is adjusted, and adaptation becomes fixed (Crawford and Chapman in Callaghan *et al.*, 1995). Subspecific genetic variation occurs in all aspects of Arctic plants such as physiology, morphology, phenology, growth form, growth rate, and reproductive development (McGraw in Chapin and Körner, 1995). However, only some of these differences can be observed in the field (Murray in Chapin and Körner, 1995) and information on population differentiation based on isoenzyme or DNA analyses is rare compared with, for example, that of temperate latitudes. In addition, the longevity of Arctic plants (Callaghan and Emanuelsson, 1985), and the long periods required for many Arctic plant species to reach a mature growth form from seed, imposes severe methodological constraints on partitioning variations in growth forms within a species between phenotypic plasticity and genetic differentiation. However, there are many examples of plant species with high levels of ecotypic differentiation, e.g., *Oxyria dygina* (photosynthetic and respiration rates), *Dryas octopetala* and *Saxifraga oppositifolia* (respiration rates and growth form), *Eriophorum vaginatum* and *Carex aquatilis* (phenology), *Phleum alpinum* (growth rates), and *Hylocomium splendens* (growth form).

Possibly as a result of high genetic and ecotypic variability, in combination with low species diversity and a lack of species to occupy available ecological niches, in the high Arctic a phenomenon exists when one and the same species dominates in different communities. Matveyeva and Chernov (in Nutall and Callaghan, 2000) call this "superdominance." Examples are the lemming *Lemmus sibiricus*, widely distributed in practically all terrestrial habitats from the very wet depressions up to summits in mountains, the crane fly *Tipula carinitrons*, and the circumpolar plants *Arctophila fulva, Betula nana sensu lato, Carex stans, Cassiope tetragona, Dryas punctata, Dupontia fisheri, Eriophorum angustifolium, E. vaginatum, Hylocomium splendens sensu lato, Aulacomnium turgidum, Drepanocladus uncinatus*, and *Ptilidium ciliare*.

Despite the wide ecological amplitudes of many Arctic species, there are still strong geographical trends associated with their distributions. At a species level, two basic types of distributions within the Arctic can be classified (Table II). Polyzonal species have wide distributions in both the tundra and the taiga and possibly elsewhere. Zonal species in the Arctic are boreal species and Arctic species. Arctic species, omitting topographical distinctions, can be subdivided into hyperarctic, euarctic, hemiarctic, and hypoarctic groups (Table II; Chernov and Matveyeva in Nutall and Callaghan, 2000). These groups represent a distributional progression from the polar deserts to the taiga boundary described later and they are described with examples in Table II.

A. Latitudinal Patterns of Diversity

Great changes in biodiversity within the Arctic often reflect sharp thermal gradients that have no analogies in other biomes. In the Siberian sector of the tundra zone over a distance of 900 km, the mean July temperature decreases from 12° to 2°C, whereas in the boreal forest belt a comparable 10°C change in mean July temperature occurs over almost 2000 km, a range across which there are three natural life zones. The gradient of summer temperature of 12°C at the treeline of the Taymyr Peninsula to 2°C in the polar desert is associated with a decrease in the number of vascular plants from approximately 250 in the south to approximately 50 in the north. Analogous decreases in plant diversity in Canada also exist (Rannie, 1986). A similar pattern occurs in the animal world, for instance, in day butter-

TABLE II

Diversity Changes with Latitude in the Arctic Region[a]

		Examples		
Category	Distribution	Plants	Birds	Mammals and invertebrates
Polyzonal	Taiga and far to the north in tundra but usually in local habitats and wet depressions	Soil algae; the mosses *Hylocomium splendens sensu lato, Aulacomnium turgidum, and Racomitrium lanuginosum;* the liverwort *Ptilidium ciliare*; the lichens *Cetraria islandica, Psora decipiens,* and *Cladina rangiferina*; the vascular species *Cardamine pratensis, Chrysosplenium alternifolium,* and *Eriophorum angustifolium*; the sedge *Carex diriscula*; the herb *Helictotrichon krylovii*; the moss *Tortula ruralis*	The common raven *Corvus corax*, the peregrine falcon *Felco peregrinus*, the white wagtail *Motacilla alba,* and the northern wheatear *Oenanthe oenanthe*	The wolf *Canis lupus* the ermine *Mustella ermine,* the weasel *M. nivalis,* the vole *Microtus gregalis,* and the mite *Chiloxanthus pilosus*
Zonal boreal	Not abundant and constrained to the south of the Arctic in benign habitats such as river valleys, south-facing slopes, and wet areas	Tree species of *Larix*; the orchid *Corallorhiza*; the shrub *Salix myrtylloides*; the sedges *Carex chordorrhiza;* the herbs *Allium schoenoprasum, Cortusa matthiolii, Galium densiflorum, Sanguisorba officianalis*; and forest mosses *Climacium dendroides, Pleurozium shreberi*, and *Rhytidiadelphus triquetrus*	The forest birds *Turdus* and *miscus* and *T. polaris* (thrushes); chiff-chaffes *Payllescopus trochilis* and *P. collybita*; and river ducks *Anas acuta, A. penelope*, and *A. crecca*	Reindeer and the wolverene, Gulo gulo
Zonal Arctic				
Hyperarctic	Polar desert and in the northernmost part of the tundra zone	Almost no plants are restricted to these zones: the following have their highest frequencies there. The grasses *Phippsia algida and Poa abbreveiata*; the herbs *Cerastium regelii, Draba oblongata, D. subcapitata, Saxifraga hyperborea*, and *S. oppositifolia*; the mosses *Dicranoweisia crispula, Bryum tortifolium, Orthothecium chryseum* and *Seligeria polaris*; and the lichens *Cetraria delisei, C. elenkenii, Dactylina ramulosa, D. madreporiformis*, and *Thamnolia subuliformis*	The dovekie *Alle alle*, the ivory gull *Pagophila eburnea, and snipe Calidris alba* and *C. maritima*	Polar bear and the collembolan *Vertaqopus brevicaudus*
Euarctic	Northern part of the tundra zone, rare in the southern part	The dwarf shrubs *Salix polaris* and *S. arctica* (this group is relatively small, but it has an important value in the subdivision of the tundra zone into subzones)	The black-bellied plover *Squatorola squatorola*, the curlew sandpiper *Calidris ferrunginea*, the snowy owl *Nyctea scandiaca*, and the snow-bunting *Plectrophenax nivalis*	The leming *Dicrostonyx torquatus* and the crane fly *Tipula carinifrons*
Hemiarctic	Throughout the tundra zone but most frequent in the middle	Most of the dominant plant species: the grasses *Arctophila fulva Dupontia fisheri*; the sedges *Carex bigelowii/arctisbirica and Carex stans*; the shrub willow *Salix reptans*, the dwarf shrubs *Dryas punctata/octopetala* and *Cassiope tetragona*; the mosses *Tomenthypnum nitens, Drepanocladus intermedius,* and *Cinclidium Arcticum*; the herbs *Lagotis minor,* and *Pedicularis hirsuta,* the moss *Polytrichum juniperinum*	The lapland longspur *Calcarius lapponicus*, the lesser golden plover *Pluvialis dominica*, and the dunlin *Calidris alpina* and *C. minuta*	The lemming *Lemmus sibiricus*, the bumblebees *Bombus hyperboreus* and *B. belteatus*, and the ground beetles *Amara alpina, Pterostichus costatus*, and *Syrphus tarsatus*
Hypoarctic	Optima in the southern tundra subzone	This group characterizes the southern tundra subzone; the shrubs *Betula nana/exilis* and sedge *Eriophorum vaginatum*	The ptarmigan *Lagopus lagopus*, the spotted redshank *Tringa erythropus*, the little bunting *Emberiza pusila*, and the bar-tailed godwit *Limossa lapponica*	The vole *Microtus middendorfi*, the ground beetle *Carabus truncaticolus*, the bumblebee *Bombus cingulatus*, and the spiders *Alopecosa hirtipes* and *Lycosa hirta*

[a] Compiled from information in Matveyeva and Chernov in Nutall and Callaghan (2000).

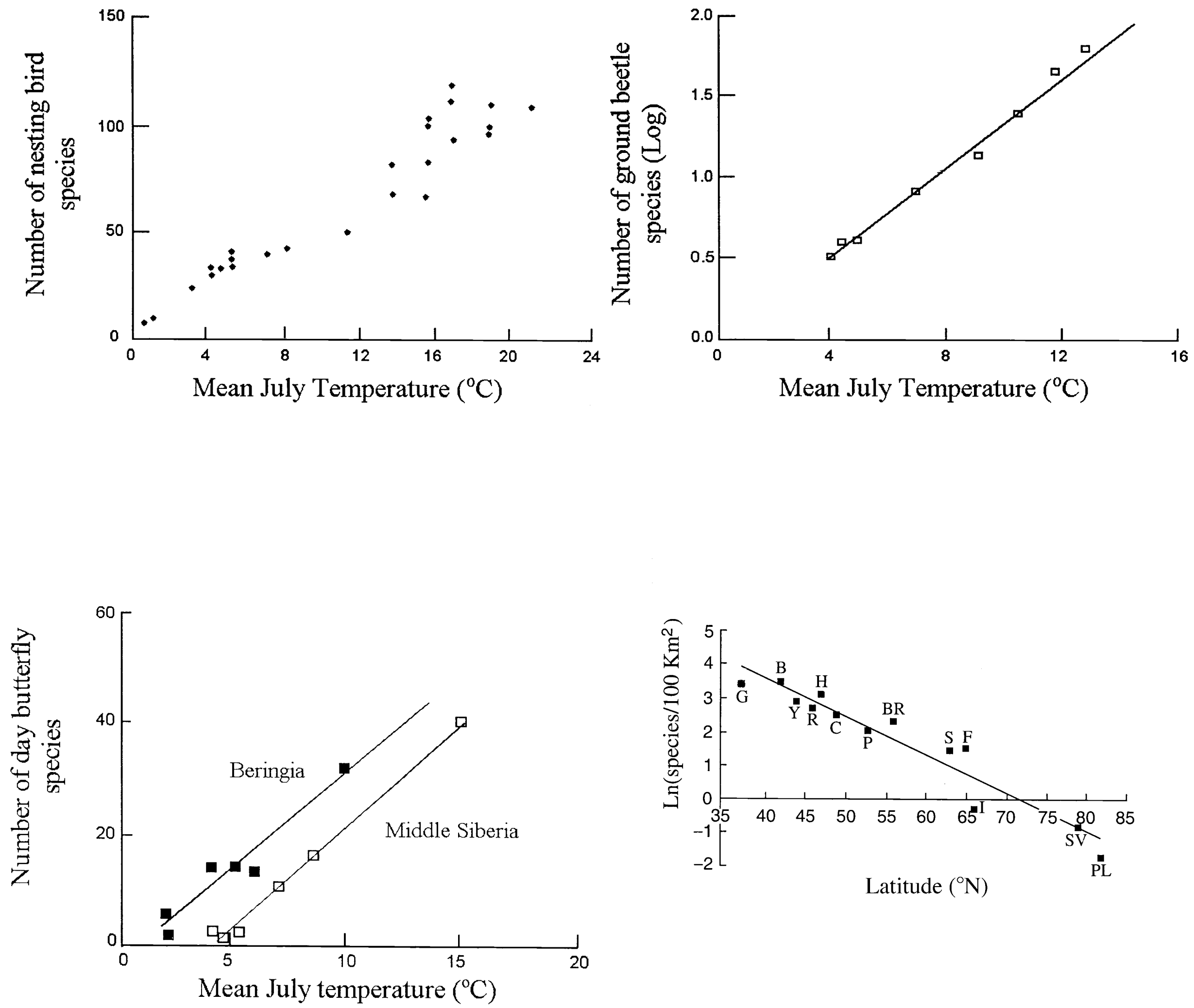


FIGURE 2 Relationships between biodiversity and/or latitude for various groups of animals and plants (after Matveyeva and Chernov in Nutall and Callaghan, 2000). Note the logarithmic scales on the right-hand graphs. G, Greece; B, Bulgaria; Y, Yugoslavia; H, Hungary; R, Rumania; C, Czechoslovakia; P, Poland; Br, Britain; S, Sweden; F, Finland; I, Iceland; SV, Svalbard; PL, Peary Land, northeast Greenland.

flies (Fig. 2) and spiders. At a larger scale, there is a general decrease in biodiversity with an increase in latitude from temperate to arctic ecosystems (Fig. 2). In this case too, decreasing biodiversity is associated with a gradient of decreasing temperature with increasing latitude, as discussed previously.

In the plant kingdom, patterns of decrease in biological diversity toward the North Pole differ among taxa. Some diminish their significance in the biota or even disappear (Ericoids), others change proportionally to the level of general diversity (Fabaceae and Rosaceae), and a third type retain a high level of biological diversity in the Arctic where their proportion becomes higher (Saxafrigaceae and Brassicaceae). Not only is there a decrease in plant species and family diversity with increasing latitude but also the paucity of life forms within the plant kingdom of the Arctic results in a simpler vertical vegetation structure than that of the forested regions further south. In the Arctic, two- or three-layered vegetation, with the height of the tallest shrub layer up to 2 m, changes into the nearly two-dimensional plant cover of polar deserts where most of the biota is concentrated into a thin film of less than 5 cm above the ground and no more than 5 cm below the ground.

Because any species belongs to a particular life (or growth) form, the changes in species composition inevitably leads to changes in the proportion of life forms in a wide sense. The proportion of rooted species (flowering plants) which control their water content to merely adnate species (cryptogams) decreases from approximately 1 : 2 at the treeline to about 1 : 5 in the polar desert. Tall shrubs are common in the south of the tundra but absent in the north. Dwarf shrubs are widely spread throughout the tundra zone but absent in polar desert ecosystems. The proportion of long rhizomatous herbs decreases while that of tap-rooted plants increases in the northern direction. Loose caespitose species are replaced by dense tufted cushions or mat-forming species (Fig. 3). Within a species, there is also a tendency to form more compact growth forms in the north compared to the south.

Latitudinal patterns of biodiversity within the Arctic may reflect global trends. Passerines and other bird taxa that are important in the tropics (e.g., kingfishers, woodpeckers, parrots, hummingbirds, and pigeons) decrease successively as a proportion of the avifauna from the equator to the poles. Passerine birds comprise about 60% of the avifauna in temperate and tropical forests, about 50% up to the northern taiga landscapes, 40% in forest–tundra, and 20% in the northern part of the tundra zone. In contrast, the proportion of Charadriiformes (40%) and geese-like birds (20%) in the Arctic avifauna is high and decreases toward the south. Charadriiformes are reduced to 20% in the north taiga and 10% in the south, and they contribute only 5% in the tropics.

B. Circumpolar Diversity

The Arctic's biota has a relatively large number and proportion of circumpolar species compared with circumboreal species. One of the obvious reasons for this is the decreasing circumference of latitudes toward the poles. Circumpolar vascular plants comprise about 45–50% of the extant flora in the south, 60–65% in the

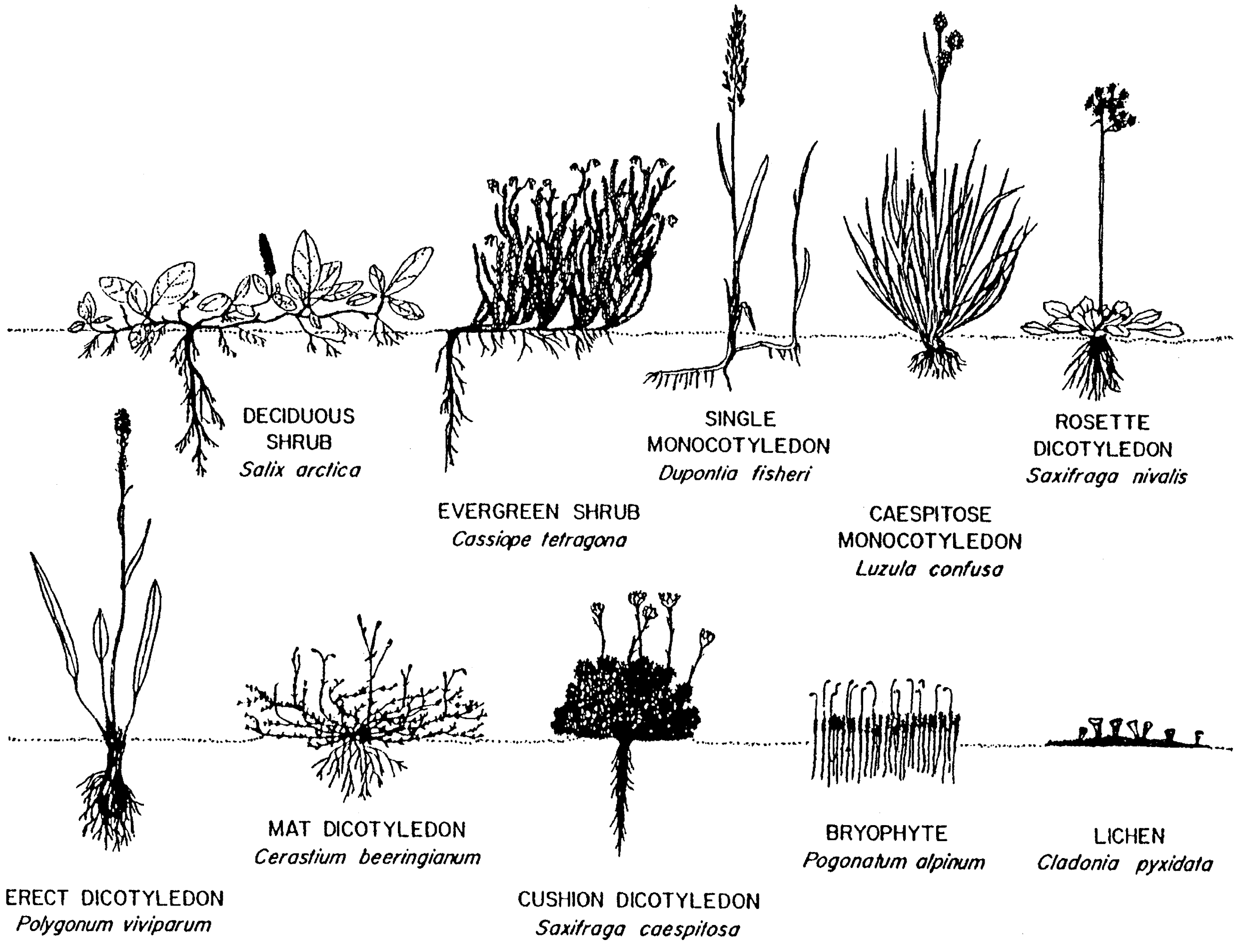


FIGURE 3 Growth forms of tundra plants, with example species. Plant functional type classifications basically follow the growth forms depicted (reproduced with permission from Webber, 1978).

north of the tundra zone, and more than 70% in the polar desert.

Less than half of the Arctic birds may be considered circumpolar because many species formally considered as circumpolar have large gaps in their distribution. Thus, about 20% of the Arctic avifauna are true circumpolar species, and only 10% of the tundra terrestrial fauna are circumpolar species. Only 2 of 50 tundra *Tipulidae* species have true circumpolar distributions. Only 1 species is conditionally circumpolar among about 40 leaf-eater beetles (Chrysomelidae) known in the Arctic. In two very common plant genera, *Saxifraga* and *Draba*, which each include about 40 species in the Arctic, only 25% (10 in each) have true circumpolar distributions. The rich genus *Pedicularis* has 6 circumpolar species out of 24, but 5 of these have large gaps in Europe, Greenland, or eastern North America. In the *Ranunculus* genus, 9 of 25 species are circumpolar. Only 1 of the 10 species of *Dryas* in the Arctic, *Dryas punctata sensu stricto*, has a truly circumpolar distribution.

Species that form monotypic genera have the most pronounced circumpolar distribution. Examples from the animal world are arctic fox (*Alopex lactopus*), snowy owl (*Nyctea scandiaca*), old squaw duck (*Clangula hyemalis*), and calliphorid (*Boreellus atriceps*) and from the plant world the grasses *Arctophila fulva, Dupontia fisheri*, and *Pleuropogon sabinei*, the herbs *Koenigia islandica* and *Oxyria dygina. Oxyria dygina* and *K. islandica* have enormous geographic ranges, being circumpolar and with extensions far south: *K. islandica* is bipolar.

Most species have distributions narrower than circumpolar. In general, there are five main types of longitudinal distribution: Species can have (i) true circumpolar distribution or, if represented on both continents, they can have (ii) amphi-Atlantic or (iii) amphi-Pacific (Beringian) distributions or they can occur only on (iv) Eurasian or on (v) North American continents. Within each continent, there are an additional five subregions, giving 10 zonations in total (Porsid, 1959): They can have a wide geographic range (Eurasian species or North American species) or occur in a more restricted area (e.g., Siberian or Alaskan species).

C. Effect of Topography on Biodiversity

The macrorelief that is represented by Arctic lowlands and mountain ranges is the main cause of interlandscape diversity and imposes altitudinal biodiversity patterns upon longitudinal and latitudinal ones. The diversity of mesorelief forms is responsible for the diversity of the vegetation at the intralandscape level. Throughout a range of tundra environments, there are recurring patterns in the distribution of diversity relative to topography. Using Russian terminology, zonal (i.e., mesic) tundra communities occupy the central environmental position within the landscape, namely, the "plakor" or watershed. These communities consist of continuous lichen–moss cover (pleurocarpous green mosses and fruticose lichens) with either a more or less sparse layer of shrubs, dwarf shrubs, sedges, and cotton grasses in the south or grasses and dicotyledonous herbs in the north. These communities have the richest species composition and are considered the true tundra communities for a given site. In addition, there is a rich diversity of intrazonal communities with fewer species and life forms developing along environmental gradients that deviate from zonal communities, including mires (without fruiticose lichens), grasslands on south-facing slopes (with only few mosses), snowbed vegetation on the north-facing slopes, dwarf shrub heaths on fell-fields (with few mosses), dense shrub thickets in deep valleys, and salt marshes.

Although different types of vegetation can be distinguished within a landscape, even greater diversity in vegetation pattern can occur over shorter distances due to microrelief (e.g., patterned ground). There are a host of forms of patterned ground (Washburn, 1956). Within a given patterned ground feature substrate stability can vary widely, influencing disturbance, soil nutrient availability, soil moisture, microclimate, and thickness of the active layer above permafrost. The relief of the patterned ground feature may influence the duration and height of the snow within a few square meters, e.g., between hummocks and hollows. The extremely localized ability of Arctic plants to influence microenvironments leads to a great complexity of plant diversity, biomass, net primary production, and soil flora and fauna characteristics with small (>10 cm) differences in relief.

Thus, despite low alpha diversity in both the plant and animal worlds in the Arctic, there is still great beta diversity both at intra- and interlandscape levels, particularly in the south. An additional resource increasing both landscape gamma and beta diversity is the presence of diversity "focal points" (Walker in Chapin and Körner, 1995) or "oases" (Edlund and Alt, 1989; Svoboda and Freedman, 1994). Moist areas below snowbeds in the Canadian polar deserts have a plant cover many orders of magnitude higher than that of the surrounding barrens and contain most of the local plant species diversity. Pingos, with their dome shape, steep slopes of diverse exposures, and gravel and sand substrate, are also "hot spots of diversity" on the flat Alaskan coastal plain dominated by tussock tundra (Walker in Chapin and Körner, 1995).

Localized topographical extremes can be associated with extreme species distribution outliers. Deep valleys with thick snow cover in winter and warm weather in summer at latitude 75°N on Taymyr have dense willow thickets of *Salix lanata* up to 2 m tall. Warm-water springs in Alaska, Chukotka, and middle-west Greenland create localized warm microclimates and many stands in the surroundings of these hot springs support extreme range disjunctions of southern, even boreal, species. Examples include stands of *Populus balsamifera* in the northern foothills of the Brooks Range and in the Chukotka Peninsula, well to the north of the tree line; ferns, orchids, and umbellifers near warm springs on Disko Island; and dwarf shrubs in some inner fjiord areas of Svalbard. Steep south-facing river banks combine warm and well-drained habitats with the possibility of migration along the river corridor and are another focal point of community diversity, often supporting trees within the tundra. Animals also create diversity hot spots, when they are dead, by providing nutrients and moisture in polar deserts and fellfields and, when they are alive, by bringing back nutrients to their nesting places or dens. Examples are bird cliff communities, fox and wolf dens, and lemming gardens. Such outliers represent potential "inocula" for recolonization (discussed later) and plant community change in a future warmer climate and support animal species with more southerly distributions. This process may have occurred about 12,000 or 13,000 years ago in Beringea when birch rapidly spread and created a "birch zone."

VI. CONTROLS OF BIODIVERSITY

A series of filters has selected, and is selecting, species and genotypes which occur in the Arctic (Walker in Chapin and Körner, 1995). The first filter is the presence of a species in the region, the second set of filters is the biogeography of species within the Arctic, and the third set of filters is internal filters within communities and the environment. Körner (in Chapin and Körner, 1995) adds a time dimension to the geographical filters. For example, the current range of species and communities in the Arctic has been determined by survivors from the thinning of past populations and floras during glacial periods together with the species which immigrated into the Arctic during the Holocene. In his alpine examples, Körner describes a second filter or sieve, grazing and freezing temperatures, and a third filter which acts at the microscale and specifically selects plant species which can tolerate certain microclimates, soil disturbance levels, and moisture regimes.

In the Arctic, during the Pleistocene, it could be argued that the first filter was environmental in that only freezing-tolerant plant species could survive in a region which became "arctic" in character. Later, during the Holocene, the next set of filters could be seen as relating to migration and dispersal abilities in that many species followed the retreating ice margins northwards, whereas some glacial relicts spread outwards from their refugia. It is possible that this filter could at least partially explain the high abundance of cryptogams, which have easily dispersed spores, relative to vascular plants in the Arctic. In general, however, it can be assumed that the main set of filters on current Arctic biodiversity are environmental rather than constraints on migration and the "available species pool." This assumption is supported because, for example, latitudinal treelines are often situated more to the south now than during the earlier Holocene (although trees never reached the high Arctic, even during the climatic optimum). Also, herbivore diversity and population sizes are limited by low primary production resulting primarily from low nutrient availability. Any local increase in nutrient availability (e.g., though animal activity) dramatically stimulates biodiversity and productivity.

The next set of filters, as suggested by Körner, act at the microscale and consist of numerous interactions among plants in communities, between plants and herbivores, and between all organisms and the microenvironment. These fine-scale filters select for organisms which are preadapted to Arctic environments or which have developed specific adaptations. Among the animals, preadaptations can be seen in traits common in boreal forest animals. In plants, preadaptations are particularly evident in those with arctic–alpine distributions (i.e., with populations in alpine areas to the south of the Arctic) and in those from nutrient-poor bogs of the temperate region. Specific adaptations of plants to the Arctic environment have been considered to be few. A constraint on adaptation rate has been the generally young nature of the Arctic flora, the longevity of many plants, and the sporadic successful completion of reproduction and seedling recruitment in areas of closed vegetation (Callaghan and Emanuelsson, 1985).

The fine-scale filters on biodiversity arising from biotic interactions were mentioned earlier. Plant competition is not thought to be a major force displacing species in the Arctic. Indeed, species removal experiments often result in little compensatory growth of remaining species. Instead, facilitation and succession are important and might explain to some extent the similarity of plant aggregations throughout large areas of the Arctic. If there were any competitive "tensions" within these aggregations, they would be expected to have broken down over time or over geographical ranges.

Herbivory is only sometimes a filter on biodiversity. Perhaps in general, herbivores do not eat particular plant species to extinction and, indeed, can stimulate plant diversity. In other cases, such as grazing by snow geese in the Hudson Bay area, high population pressure can lead to ecological cascades ending with loss of plant populations and ultimately soil erosion. Similar dramatic impacts on vegetation can occur when population peaks of two different herbivores, both feeding on the same plant, occur within a short period of time. Analogs of tundra vegetation in Fennoscandia can be produced when subarctic birch forests are defoliated by insects and regenerating trees are browsed by reindeer. Also, browsers such as the megafauna of the Pleistocene and early Holocene had the ability to maintain open vegetation (Zhimov *et al.* in Chapin and Körner, 1995), whereas grazers can control dominance in plant communities, e.g., by suppressing moss growth. Although such impacts can be dramatic at the local scale, the generally wide distributions of many Arctic species ensure survival of the species. Survival of rare species and ecotypes, however, has a greater risk.

V. CONSEQUENCES OF BIODIVERSITY

A. Environmental Consequences of Biodiversity

One way in which we can readily consider the importance of biodiversity within processes is to understand the potential impact of biodiversity decline. Primary successional processes are indicative of the potential impact of biodiversity loss. One of the most thoroughly studied successional series within the Arctic is that present at Glacier Bay, Alaska (Chapin *et al.*, 1994). The retreat of the glaciers during the past 200 years has led to the development of a series of communities the spatial arrangement of which can be considered analogous to a sequence through time. At the foot of retreating glaciers is a harsh environment of exposed glacial till that lacks soil organic matter and soil nutrients, and is a potential zone of drought. Mats of nitrogen-fixing cyanobacteria, lichens, and liverworts, with scattered forbs (e.g., the nitrogen-fixing *Dryas drummondii*), colonize these environments. The process of pedogenesis, involving the accumulation of organic matter and plant-available soil nutrients, then proceeds and the succession sequence passes through many stages from alder woodland (*Alnus sinuata*) to sitka spruce (*Picea sitchensis*) forest. The seedlings of key species from each successional stage at Glacier Bay were found to germinate more readily within the preceding successional stage than within that of their own (Chapin *et al.*, 1994). Every step of the successional process enhances the development of species dominant within the next stage. Given the loss of any individual step within this series, the composition of the community may develop along an entirely different trajectory, or the environment may never develop beyond a given point.

The nature of arctic environments is such that events leading to the initiation of primary succession, e.g., ice-based processes such as frost heave, avalanche, and debris flows and high runoff due to permafrost-impeded drainage, are common. For any site undergoing primary succession, the availability of propagules of primary colonizers within that area is essential. An underlying level of diversity is necessary throughout the environment for appropriate colonizer species to be within range of disturbance events. For example, in the Alaskan tundra, gravel pads and borrow pits are often initially colonized by species from nearby riparian systems. Without habitat diversity at the landscape level these colonizing species would be absent. Therefore, both the distribution and the absolute number of organisms are important for ecosystem function. As discussed earlier, organisms may maintain the habitat diversity of arctic systems. The activity of burrowing animals such as ground squirrels or arctic foxes may provide exposed areas of disturbed soil that act as potential refugia for plant primary colonizers. This is an example of faunal diversity promoting floral diversity.

As succession proceeds, the potential for interaction between organisms increases because their zones of influence are more likely to overlap with increasing organism density (this is especially true of plants). Because of an increase in the number of species, the potential complexity of interactions within a community also increases. Therefore, there may be a parallel increase in the level of biodiversity and the degree and complexity of interactions between individuals. This type of pattern might be found along the latitudinal gradient from high Arctic polar semidesert toward more productive tundra environments.

Types of interaction include direct resource-based interactions (e.g., competition) and indirect interactions, possibly involving higher trophic levels. For example, there is evidence that herbivores may encourage the development of vascular plants by preventing the formation of dense moss mats in certain tundra environments (Zhimov *et al.* in Chapin and Körner, 1995; R. van der Wal, personal communication). Because of the mobile nature of larger herbivores (e.g., reindeer and musk ox), their importance in the functioning of a community at any given time may not be clear. As in the case of plant primary colonizers, the key role that they play, and their importance in terms of maintaining diversity at both the species and the community level, may only be apparent at particular points in time or space.

Facilitative interactions may be of particular importance within Arctic communities. In the case of plant interactions, facilitation may act through the amelioration of environmental severity and may occur within all successional stages, including mature communities. Facilitative plant species may be keystone species within an environment; for example, on areas of frost heave-disturbed scree the growth of spreading plant species with persistent nets of rhizomes or roots has been shown to stabilize the scree and to allow the development of other plant species within the stabilized zone. A critical factor in this case is the morphology of the colonizing plant. Other types of plant growth form may have different facilitative effects. Although all plants may provide shelter to their neighbors, there are certain species and growth forms, especially cushion plants, which can develop in the absence of this initial shelter and can raise temperature differentials between their meristems and the air by more than 25°C. They subsequently provide shelter for others. In one alpine example, five species of cushion plant, including *Silene acaulis*, together "hosted" 93 other species of plants. In another example, grasses growing in moss cushions in the high Arctic grew dramatically faster than those growing outside moss cushions (Fig. 4). Diversity of form (Fig. 3) is therefore perhaps one aspect of Arctic biodiversity that may be critical for ecosystem function.

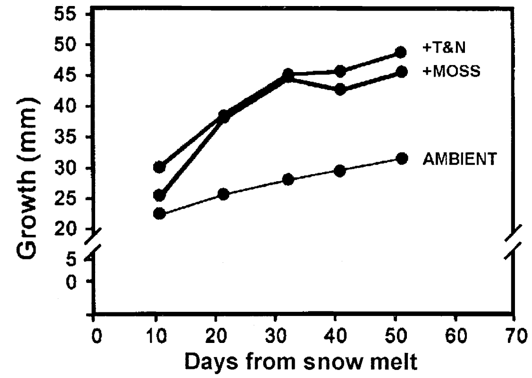


FIGURE 4 Impact of mosses in determining the growth of vascular plants in the high Arctic (the example is the grass *Poa arctica* growing in cushions of the moss *Drepanocladus uncinatus*). The impact of the moss on the growth of the grass, when compared with the grass growing outside moss cushions, is equivalent to growing the grass under plastic greenhouses with nutrient addition (+T and N) (after Press *et al.*, 1998).

Differences in plant growth form also strongly affect the movement of resources within an environment. Spreading rhizomatous species that maintain physiological connections between individuals may move nutrient resources away from localized resource-rich areas. Species with spreading root systems but a localized canopy may in turn concentrate resources within a given area. Similarly, differences in the behavior of animals may also lead to variability in resource distribution and resource movement throughout the environment. During the winter period when their grazing is confined to roots, rhizomes, and the stem bases, lemmings can accumulate within their nesting areas small reservoirs of nutrients in plant tissue and droppings. Other grazing species such as geese may redistribute nutrients throughout the environment.

Not only are nutrient distribution processes dependent on the variability in life form, and hence diversity, but also nutrient cycling and soil conditions (e.g., temperature, moisture, and pH) may be strongly related to the diversity of organisms within an ecosystem. Hobbie (in Chapin and Körner, 1995) describes how differences in plant traits may directly and indirectly influence biogeochemical processes within tundra plant communities. For example, species with high productivity rates (generally deciduous and graminoid species) may increase the rate of nutrient turnover, woody species may increase the sequestration of carbon, and graminoid species may attract herbivores and increase the degree of herbivore nutrient cycling. Interspecific variability in rooting depth and leaf area influence soil moisture conditions, which in turn may influence soil nutrient availability, soil biota diversity, ion uptake, and soil pH. Leaf litter tissue quality differences may also influence soil microbial processes and decomposition rates. Mixtures of litter from more than one arctic plant growth form, particularly herbs and dwarf shrubs, have been shown to decompose at a faster rate than single-species litters (H. Quested *et al.*, personal communication). The diversity of a plant community will therefore have important consequences for soil environmental conditions and terrestrial nutrient cycling which is a major constraint on productivity of Arctic ecosystems.

Species diversity may also affect atmospheric environmental conditions by influencing soil–atmosphere exchange processes. Verville *et al.* (1998) found that sedges contribute to methane emission from Arctic wet meadow communities by acting as a conduit for methane release from anaerobic soils (rather than by producing methane directly). In contrast, mosses tended to limit the evolution of methane (Fig. 5).

B. Natural Resources and Environmental Exploitation

At the most simplistic level, biodiversity can be considered a natural resource. Arctic ecosystem biodiversity, in terms of species number (i.e., alpha diversity), is not high. One key future of Arctic biodiversity as a resource is its unique components, i.e., species (and processes) that are confined to Arctic ecosystems. However, existence *per se* is not commonly considered to be a resource, and we tend to regard resources as items for exploitation by ourselves, either currently or in the future.

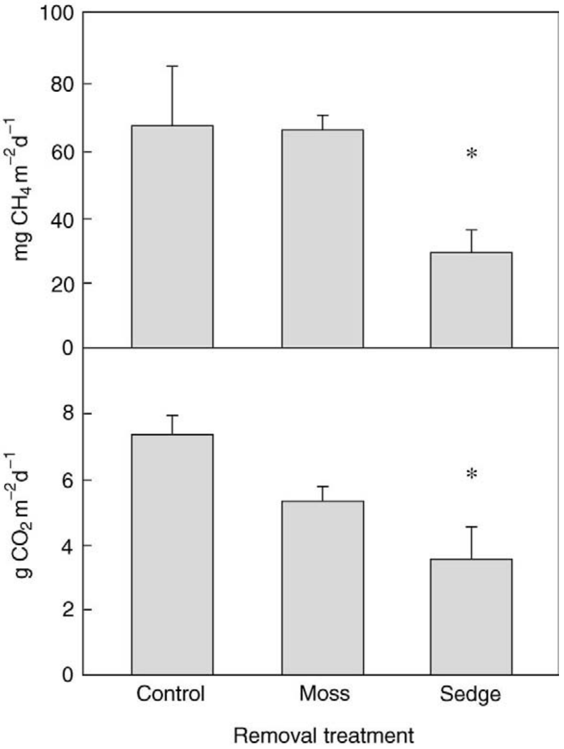


FIGURE 5 Methane and carbon dioxide fluxes in short-term species removal plots in wet meadow tundra. Sedges acted as a conduit for methane and carbon dioxide uptake: Mosses did not. * Statistically significant differences among species-removel treatments ($p = 0.05$). (Reproduced with permission from Verille *et al.*, 1998.)

Certain types of human exploitation of Arctic natural resources are limited by biodiversity and by productivity, namely, those in which an organic crop is removed from the environment. Biodiversity determines the number of resources available and productivity limits the rate at which they can be exploited. For humans to live within the Arctic using extant natural resources, they must travel over a wide area to obtain and concentrate sufficient energy, nutrients, and raw materials. The Saami people of northern Fennoscandia follow their reindeer because productivity in general is very low and it is necessary for the reindeer herds to move over a vast area in order to obtain sufficient resources. This is true both for the semidomesticated reindeer herds of the Saami and for the wild caribou herds of North America. One impact of low biodiversity in this context may be the necessity of exploiting one species intensively, such as the Saami exploitation of reindeer, rather than being able to switch exploitation between species on a seasonal basis.

Nonindigenous human exploitation of wildlife has involved hunting to accumulate raw material for manufacturing processes (e.g., furs and blubber). The replacement of many animal-derived products with synthetic materials derived from oil has reduced the pressure on Arctic animal populations. However, hunting of Arctic wildlife still occurs, either by indigenous peoples, such as the Saami or Inuit, or for sport. This latter type of hunting is just one particular branch of the development of tourism within the Arctic. Ecotourism depends heavily on the wildlife of the Arctic. The biodiversity of the Arctic, including unique Arctic species, and the paucity of species and barren nature of high-latitude Arctic systems in particular (i.e., their inherently low biodiversity) may be features of Arctic environments that are attractive to visitors. In this case, biodiversity can be an exploitable resource, and as ecotourism develops perhaps the monetary value attached to the biodiversity of the Arctic will increase.

A more intangible natural resource dependent on Arctic biodiversity is the potential for future acquisition of knowledge. The low biodiversity and process simplicity within Arctic ecosystems makes them a valuable resource for testing ecological theories. Arctic ecosystems may provide a second type of scientific resource, i.e., they might contain undiscovered or unrecognized useful natural substances. Although the low biodiversity of Arctic ecosystems suggests that the number of possibly useful substances is less than that in diverse ecosystems (e.g., rain forest), we may find unique substances that have evolved in response to the Arctic environment. For example, it may be that the Arctic provides us with a source of natural antifreeze chemicals. Similarly, the mechanism whereby some Arctic species can tolerate extreme anoxia might be important in understanding the problems associated with the susceptibility of some trees and crops to waterlogging.

Mineral resources are another natural product extracted from Arctic environments. However, in this case the problem of exploitation is not one of low biodiversity but rather that of exploiting the resources without damage to biodiversity, as discussed in the following section.

VI. THREATS TO BIODIVERSITY IN THE ARCTIC

The Arctic is currently undergoing rapid changes: cultural and sociological changes resulting in exploitation of natural and nonrenewable resources, changes in climate, and changes in pollution—both local and transboundary (Nutall and Callaghan, 2000). Many of these changes are predicted to accelerate in the future and these have implications for changes in biodiversity within the Arctic.

Extractive industries (oil, nickel, etc.) have resulted in large local impacts on species, including reductions in the diversity of soil microorganisms (Evdokimova in Callaghan *et al.*, 1995). Examples of impacts of mineral extraction are found in towns of the Kola Peninsula, but there has been surprising resilience of the vegetation and pollution-tolerant forms are being actively selected for (M. Kozlov, personal communication). Oil spills in Alaska and the Perchora region of Russia have had dramatic effects on local populations of both plants and animals. However, the large and widespread distributions of Arctic biota provide a system in which changes in biodiversity from this source over wide areas are unlikely. More serious is the slow recovery potential of Arctic vegetation which exacerbates disturbances to ecosystems.

Harvesting of natural resources on land in the Arctic consists of hunting, collecting (e.g., berries), and reindeer husbandry. Effects of hunting have probably had serious impacts on a small proportion of Arctic biota over the Holocene. It is likely that the large northern grazers and herbivores such as the mammoth were hunted to extinction at the beginning of the Holocene, whereas hunting within the past 300 years probably hastened the demise of the great auk. Following the advent of firearms in the north, wolf populations have been reduced to just a few animals in the subarctic of Fennoscandia. Currently, international agreements protect some species such as polar bear and their populations have increased. However, the health of many species of birds which migrate to the Arctic is determined more by hunting practices in the overwintering grounds than in the Arctic: This is more difficult to legislate.

Changes in reindeer husbandry practices in Fennoscandia during the past 300–400 years from hunting to herding and, recently, from herding to almost farming in some areas have had impacts on predators (hunted by husbanders) over large areas and on grazing pasture productivity and biodiversity. Restrictions on the nomadic lifestyles of reindeer husbanders both in Fennoscandia (due, for example, to constraints of national boundaries) and in areas of the Russian Arctic (due to imposition of a sedentary lifestyle for women and children) have resulted in large-scale impacts on vegetation with an often dramatic reduction in lichen cover. Changes in lichen biodiversity remain to be documented but are likely to be pronounced.

Pollution in the Arctic is still generally less than in more southerly regions and the pathways of pollutants into the Arctic and their concentrations within Arctic biota have been well documented recently (Reiersen in Nutuall and Callaghan, 2000). However, little is known about the sensitivity of Arctic biota to the wide range of pollutants which are found there.

During the 1.8 million years of the Pleistocene, the Arctic has undergone severe climatic fluctuations associated with glacial and interglacial periods. Even within the last interglacial period of the Holocene, climate has changed greatly, resulting in a current cool period which is associated with lower and more southerly treelines in many places in the Arctic than found in the early Holocene. Since the Little Ice Age, however, which ended in the north Atlantic region about 150 years ago, mean annual temperatures have been increasing. Since at least 1960, temperatures have been changing throughout much of the Arctic. In continental Alaska and eastern Canada, and also in central Siberia, mean annual temperatures have risen by up to 1.5°C per decade. However, in eastern Canada and midwest Greenland, temperatures have been decreasing by the same amount. Overall, though, the Arctic's temperature has increased and there has been a concurrent increase in the "greenness index" of vegetation in the north, as seen by satellite images, which suggests an increase in the growing season of approximately 11 days over the past 20 years. The recent increases in temperature in the Arctic have been associated with increased pest outbreaks in the taiga, increased frequency of forest fires, and a warming of the permafrost with an increase in disturbance events.

General circulation models which predict future climates for various concentrations of greenhouse gases all agree that future warming will be greatest in the Arctic, and particularly during wintertime. Because of the strong correlations between biodiversity and temperature discussed previously, it is expected that future warming will have a large impact on biodiversity in the Arctic. Experiments in a range of Arctic habitats which simulate warmer conditions show that one of the first effects of warming on plant communities is a change in the dominance of vascular plant species existing when the manipulations started and an increase in biomass (Chapin *et al.*, 1995; Press *et al.*, 1998). However, as the vascular plants increase in biomass, lichen biomass and cover decrease and some moss species (e.g., *Hylocomium splendens)* also decrease in abundance. In the case of lichens, the decreases seen in the experiments parallel decreases found along natural geographical gradients of temperature. The major implications of the experiments are therefore that, over the medium term, warming will change the dominance among vascular plant species, the abundance of lichens and mosses will decrease, and the immigration of more southerly species will not happen. Over the longer term, clearly biodiversity of vascular plants will increase but that of cryptograms will decrease. In addition to this decrease in lichens is the impact of changing reindeer husbandry discussed previously. Overall, the Arctic's important role as a reserve of primitive plants is threatened, and because there is such a great cover of lichens in the Arctic, this will have an effect on their biodiversity at a global level.

In contrast to predicted decreases in biodiversity of the mosses and lichens, warming may increase diversity of all plant groups in some places. Glacier forefield expansion as a result of glacier retreat during warming and the open ground of polar deserts and semideserts offer open niches for the establishment of new plants. Because colonization of these habitats is usually by seeds and spores, colonization can expand the local species and subspecific genetic diversity. This is in marked contrast to closed areas of vegetation in the mid- and subarctic where plant establishment is mainly via clonal growth. However, even here, increases in disturbance due to thawing permafrost are likely to open up niches and allow new genets to establish.

VII. CONCLUSION

Although biodiversity in the Arctic is low compared with that in other regions, a surprisingly varied array of landscapes, ecosystems, species, ecotypes, and topographic micropatterns exist. Relationships between diversity and environment are superficially simple because land-use impacts are relatively few in the Arctic. However, beneath the surface, interactions between the low temperatures of the Arctic and its biota are complex: Temperature has many direct effects on the biota but also controls nutrient cycling in cold soils which severely constrains the abundance of vascular plant species and the organisms which depend on them. Importantly, interactions between plants and climate are two-directional: Plant growth forms facilitate dramatic differentials to occur between plant and air temperature, whereas tundra soils have sequestered the greenhouse gas carbon dioxide throughout much of the Arctic. Currently, rapid sociological and environmental changes are occurring in the Arctic. Although low biodiversity is compensated for to some extent by large and widespread populations, the vast wilderness areas are nevertheless under numerous threats. Our major challenge is to further document (Conservation of Arctic Flora and Fauna, 1996) and monitor the biota in the remote Arctic wildernesses and to understand both how biodiversity patterns are formed and how they contribute to the functioning of ecosystems. Only then can we hope to establish conservation strategies that can cope with rapid environmental change.

Acknowledgments

We gratefully acknowledge Dr. C. Jonasson, D. Hollingworth, and C. Aune, who kindly helped with the preparation of some of the figures.

See Also the Following Articles

ANTARCTIC ECOSYSTEMS • BOREAL FOREST ECOSYSTEMS • LATITUDE, COMMON TRENDS WITHIN

Bibliography

Aleksandrova, V. D. (1980). *The Arctic and Antarctic: Their Division into Geobotanical Areas.* Cambridge Univ. Press, Cambridge, UK.

Bliss, L. C., and Matveyeva, N. V. (1992). Circumpolar arctic vegetation. In *Arctic Ecosystems in a Changing Climate* (F. S. Chapin, III, *et al.*, Eds.), pp. 59–89. Academic Press, San Diego.

Callaghan, T. V., and Emanuelsson, U. (1985). Population structure and processes of tundra plants and vegetation. In *The Population*

Structure of Vegetation (J. White, Ed.), pp. 399–439. Junk, Dordrecht.

Callaghan, T. V., Oechel, W. C., Gilmanov, T., Molau, U., Maxwell, B., Tyson, M. J., Sveinbjörnsson, B., and Holten, J. I. (Eds.) (1995). Global change and arctic terrestrial ecosystems. Proceedings of papers contributed to the international conference, Oppdal, Norway, August 21–26. European Commission Ecosystems Research Report No. 10. European Commission Luxembourg.

Chapin, F. S., III, and Körner, Ch. (Eds.) (1995). Arctic and alpine biodiversity: Patterns, causes, and ecosystem consequences. *Ecol. Stud.* **113**.

Chapin, F. S., III, Walker, L. R., Fastie, C. L., and Sharman, L. C. (1994). Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. *Ecol. Monogr.* **64**(2), 149–175.

Chapin, F. S., III, Shaver, G. R., Giblin, A. E., Nadelhoffer, K. J., and Laundre, J. A. (1995). Responses of arctic tundra to experimental and observed changes in climate. *Ecology* **76**, 694–711.

Conservation of Arctic Flora and Fauna (1996). Atlas of rare circumpolar endemic arctic vascular plants (draft). U.S. Fish and Wildlife Service, Anchorage, Alaska.

Edlund, S. A., and Alt, B. T. (1989). Regional congruence of vegetation and summer climate patterns in the Queen Elisabeth Islands, Northwest territories, Canada. *Arctic* **42**, 3–23.

Nutall, M., and Callaghan T. V. (2000) *The Arctic: Environment, People, Policy*. Harwood, Reading, UK.

Porsid, N. (1959). *Circumpolar Arctic Flora.* Clarendon, Oxford.

Press, M. C., Callaghan, T. V., and Lee, J. A. (1998). How will European arctic ecosystems respond to projected global environmental change? *Ambio* **27**, 306–311.

Rannie, W. F. (1986). Summer air temperature and number of vascular species in arctic Canada. *Arctic* **39**, 133–137.

Svoboda, J, and Freedman, B. (Eds.) (1994). *Ecology of a Polar Oasis. Alexandra Fjord, Ellesmer Island, Canada.* Captus Univ. Publications, Toronto.

Verville, J. H., Hobbie, S. E., Chapin, F. S., III and Hooper, D. U. (1998). Response of tundra CH_4 and CO_2 flux to manipulation of temperature and vegetation. *Biogeochemistry* **43**(3), 215–235.

Washburn, A. L. (1956). Classification of patterned ground and review of suggested origins. *Geol. Soc. Am.* **67**, 823–866.

Webber, P. J. (1978). Spatial and temporal variation of the vegetation and its productivity. In *Vegetation and Production Ecology of an Alaskan Arctic Tundra* (L. L. Tieszen, Ed.), pp. 37–112. Springer, New York.

Yurtsev, B. (1997). Effect of climate change on biodiversity of arctic plants. *Ecol. Stud.* **124**, 229–244.

ARTHROPODS (TERRESTRIAL), AMAZONIAN

Joachim Adis
Max-Planck-Institute for Limnology, in cooperation with National Institute for Amazonian Research

GLOSSARY

arboricolous Tree inhabitants ("arboreal").
assemblage Collective occurrence of several individuals representing the same species or several species without interspecific relationships.
bionomics Mode of life of a species.
biotope Characteristic living space, of distinguishable nature compared to its environment, of a (bio)community [also "(bio)coenosis"].
bivoltine Species with two generations per year.
community Collective occurrence of individuals representing several species with, at least in part, interspecific relationships (also "coenosis").
flood resistance Submersion ability of weeks or months.
flood tolerance Submersion ability of few hours up to several days.
habitat Characteristic living space or site of a species.
plastron Thin layer of air, held by specific body structures, into which oxygen from the surrounding water is added by means of diffusion to the same extent as oxygen is withdrawn by breathing.
plurivoltine Species with several generations per year (also "multivoltine").
quiescence Dormancy in which inhibition of development depends directly on environmental factors.
terricolous Soil inhabitants.
Univoltine (Bivoltine) Species with one (two) generations per year.

ESTIMATION OF THE biodiversity in terrestrial arthropod taxa of the "bioregion" Amazonia depends on solid data of species, collected from different biotopes and habitats, that include comparative long-term information on their bionomics and ecology. Some examples, mostly deriving from studies in Central Amazonia near Manaus, are available and clearly show the difficulties in achieving a more general statement.

I. TAXA AND NUMBER OF SPECIES

Approximately 2% of the world's Arachnida and 3% of the Myriapoda live in Amazonia. However, Ricinulei represent 28%, Schizomida and Scolopendromorpha 9%, and Pauropoda more than 7% of species known worldwide in the respective taxa (Table I). To estimate the extant species in Amazonia is difficult, especially in those cases in which numbers differ by two or three different orders of magnitude from the species described worldwide. For mites which, like spiders, are considered a mega- or hyper-diverse taxon in the Arachnida, opinions of taxonomists differ substantially. This is also due to different methods used to derive such estimates (Platnick, 1999). In millipedes, approximately 250 species have been described from the Amazon drainage basin but 5000 are estimated to exist (Table I). This is based on the assumption that most genera originated in the Andes, where most of the known species now occur but there have been few sampling efforts.

Mega-diversity and high densities are attributed to Neotropical insects, in particular those inhabiting forest canopies. The dominant taxa in the canopy of non-flooded primary forests in Central Amazonia are Formicidae and Diptera. In studies by Guerrero in 1995 and 1996 near Manaus (Reserva Ducke), these two taxa represented 52 and 10%, respectively, of the 325 arthropods/m^2 obtained on average by fogging the canopy of 40 trees (nine species of the families Sapotaceae and Lecythidaceae) once each during the dry and rainy seasons. In another study in this reserve, two canopies of a widely distributed Amazonian tree, *Goupia glabra* Aubl. (Celastraceae; height 38 and 45 m) were fogged in intervals of 6 or 24 months (1991–1994). A total of 95 ant species were found on a single tree. In comparison, there are 105 ants species recorded from all of Germany. Of the 124 ant species obtained from both trees, more than one-third were represented by singletons, i.e., one specimen per species. Data also indicated a biotic interaction between predatory ants (probably *Crematogaster* spp.) and the gall-building Cecidomyiidae and between Cecidomyiidae and the parasitic Hymenoptera. This was not determined from previous studies in the tropical canopy (Adis *et al.*, 1998a).

Comparative to the high density and biodiversity of Formicidae and Diptera in tree canopies is that of Acari and Collembola on the forest floor. These two taxa represent between 50 and 80% of the 26,000–74,000 terricolous arthropods extracted per square meter from 14-cm soil depth in non-flooded "terra firme" forests of Central Amazonia. However, due to taxonomical difficulties in both groups, data on species diversity are scarce and based on adult animals, although just as many immatures have been caught. The highest number of oribatid mites, 97 morphospecies, were reported by Wunderle from a primary forest near Pucallpa, Peru. In the Manaus area, Franklin found 71 and 74 morphospecies of oribatid mites in a primary forest on yellow latosol and on white sand, respectively, and 57 morphospecies in a 3-year-old secondary forest on yellow latosol (previous primary forest that was cut and burned). The ratio between the total number of specimens and morphospecies decreased accordingly, being 53.3, 36.3, and 22.8, respectively. In the same forests, Oliveira obtained 74, 65, and 65 morphospecies, respectively, of Collembola (53–64% represented Isotomidae and 7–16% Entomobryidae), but specimens : morphospecies ratio was lowest in the primary forest on white sand (49.5) compared to the primary and secondary forests on yellow latosol (67.6 and 61.8, respectively). Statistical analysis of data indicated that morphospecies similarity is dependent on both the nature of the soil (in springtails) and human disturbance (in oribatid mites). Only 32 oribatid morphospecies were obtained from pastures near Manaus.

II. ORIGIN AND GEOGRAPHICAL DISTRIBUTION OF SPECIES

Little is known about the centers of evolution and dispersal for arthropod taxa which occur in Amazonia. Several pathways are currently being discussed.

A. Origin in the Neotropics with a Generally Wide Range of Geographical Distribution (Chilopoda: Scolopendromorpha)

No scolopendromorph taxon of generic level is known as Amazonian endemic as it is at least present in other areas of the Neotropical realm. Studies of Schileyko showed that only 25% of the total 52 species currently known in Amazonia, representing 6 of the total 11 genera, seem to be endemic. The wide geographical distribution might be related to the relatively large size, high mobility, swimming ability, and predaceous mode of life of an ancient group.

TABLE I

Families, Genera, and Species of Taxa Representing the Arachnida and Myriapoda in the World and in Amazonia (up to 1999) as Well as Estimates of Existing Species (Fossil Taxa Not Included)[a]

				Species	
		Families	Genera	Described	Estimated
Arachnida	World	>566	>8869	>92,529	
	Amazonia	>136	>480	>1592	
Araneae	World	106	±3200	±37,000	76,000–170,000
	Amazonia	>65	>300	>1000	4000–8000
Palpigradi	World	2	6	80	100
	Amazonia	1	1	1	
Uropygi (Thelyphonida)	World	2	16	101	
	Amazonia	2	2	3	10
Amblypygi	World	5	20	126	
	Amazonia	1	2	11	
Schizomida	World	2	31	195(219)	
	Amazonia	1	3	10	
Solifugae	World	12	153	1065	(1115)
	Amazonia	1	1	1(2)	
Pseudoscorpiones	World	24	430	3100	3500–5000
	Amazonia	12	31	75	>150
Ricinulei	World	1	3	53	85
	Amazonia	1	1	15	25
Opiliones	World	44	>1554	>4559	
	Amazonia	13	75	160	
Cyphophthalmi	World	6	28	109	
	Amazonia	3	4	4	
Palpatores	World	15	>250	1000–2000	
	Amazonia	1	5	20	
Laniatores	World	23	1276	3460	
	Amazonia	9	66	136	
Scorpiones	World	18–20	156	1250–1500	6000–7500
	Amazonia	4	12–14	68–111	200
Acari	World	350–422	3300–4000	45,000	0.5–1 million
	Amazonia	35	50	150–300	20,000–250,000
Myriapoda	World	158 (159)	2167	>15,096	59,400–92,500
	Amazonia	28 (29)	>94	>423	5600–7600
Chilopoda	World	20 (21)	325	>3196	6850–6950
	Amazonia	8 (9)	26	>115	400
Scutigeromorpha	World	1	16	>80	100–150
	Amazonia	1	1	2	5
Lithobiomorpha	World	2	95	1500	2000
	Amazonia	1	1	2	5
Craterostigmomorpha	World	1	1	1	?5
	Amazonia	0	0	0	0
Scolopendromorpha	World	3	33(32)	515–616	700–800
	Amazonia	2	11	52	90
Geophilomorpha	World	13 (14)	180	1100	4000
	Amazonia	4 (5)	13	60	300
Diplopoda	World	131	1800	11,000	50,000–80,000
	Amazonia	16	55–60	250	5000–7000
Pauropoda	World	5	29	700	2000–5000
	Amazonia	2	8	52	>200
Symphyla	World	2	13	200	?500
	Amazonia	2	4	5	10–20

Note. Numbers in parentheses include forthcoming descriptions and revisions.

[a] Modified with permission from Adis and Harvey (2000).

B. Origin in the Andes and/or the Guyanian Shield and Subsequent Dispersal into the Amazon Basin (Archaeognatha: Meinertellidae; Coleoptera: Carabidae; Diplopoda: Polydesmida)

Representatives of the Meintertellidae probably advanced from the Andes into the lower situated forests of the Amazon Basin. During this process, the formerly petrophilous animals, inhabiting mountain floors, adapted to an arboricolous mode of life in the trunks and canopies of trees. They even acquired a vertical jumping ability on the tree trunk, which was found for the first time in the arboricolus species *Neomachilellus adisi* in a black-water inundation forest of Central Amazonia. In these periodically flooded forests the Meinertellidae succeeded to colonize the forest floor as well. Flood resistance of eggs (observed in *N. scandens*) and the determination of egg development by the flood pulse (= quiescence duration correlated with the period of inundation) are prerequisites for a potential distribution on the waterway (on logs or floating meadows). *Neomachilellus adisi* was also found in an inundation forest above Leticia (Columbia) and might have reached Central Amazonia on the Rio Solimões-Amazonas. Similarly, the *Neomachilellus* species which currently inhabit the Caribbean and the coastal regions of the United States might originally have been transported by ocean currents from the mouth of the Rio Amazonas and the Rio Orinoco to these areas. An actually waterproof, resistant blastoderm cuticula and a presumably long duration of egg development during transportation (1 year or more, as in European species) certainly favor a geographical distribution on waterways. An interruption of the quiescence due to drying of the means of transport used, a rapid eclosion of juvenile animals, and a fast achievement of maturity (all observed in *N. scandens*) favor the new colonization of a biotope as well. Useful for this colonization is the swimming ability and a temporal survival on the water surface (flood tolerance), which was reported for several species of the Machiloidea (Adis, 1992; Sturm, 1984).

Erwin and Pogue presented biogeographical maps with centers of endemism, based on taxa distribution and cladograms, for sister groups of the arboricolous carabid beetles genus *Agra* from Neotropical forests. Patterns of species evolution are directed, in part, from the North Atlantic coast (including the northern Guyanian shield) to upper Amazonia (*palmata* lineage) and from the North Atlantic coast to either upper and middle Amazonia or lower Amazonia (*erythropus* lineage). According to Erwin (1991), these hypothetical "centers and corridors of radiation" can be discovered and targeted for long-term protection through analyses of diverse arthropod groups (beetles in particular) and detection of congruent patterns among radiation lineages.

C. Origin in Non-flooded and/or Floodplain Forests of the Amazon Basin (Scorpiones, Chilopoda: Geophilomorpha)

Many scorpion species apparently require strict environmental conditions and thus are found only in narrow zones of distributions. Very few species currently known from eastern Amazonia are present in western Amazonia. One of the four principal "corridors of distribution" (or radiations) in South America is located in Central Amazonia. Rupture zones and eco-barriers between these corridors are attributed to geo-historical events. For instance, the rupture zone of distribution found in the region of the Orinoco delta might be due to the proposed existence of a large inland lake in Central Amazonia during the late Pleistocene and early Holocene (Lourenço, 1994).

In the Geophilomorpha, half of the approximately 100 Neotropical species representing the genera *Schendylops*, *Pectiniungis*, and *Ribautia* seem to be endemic to areas identified as Pleistocene forest refuges (see Section III) based on evidence from woody angiosperms (Prance, 1982). Valleys of the major rivers (Amazon, Orinoco, Paraná, and Paraguay) are suggested to be the main routes of subsequent distribution, even to high-altitude sites (Pereira *et al.*, 1997)

D. Origin in Treeless Floodplains of the Amazon Basin (Coleoptera: Carabidae)

According to Erwin, the main centers of evolution in carabid beetles are open floodplains along rivers or lakeshores in the equatorial tropics with somewhat stable conditions. Taxon pulses which effect these flying and highly vagile waterside generalists caused, by means of adaptations, their dispersal in three directions: (i) colonization of tree-tops (canopy specialists), (ii) colonization of mountains (altitude specialists), and (ii) colonization of temperate regions (climate specialists). For the colonization of temperate regions, seasonality is considered a basic requirement. Seasonal changes, primarily due to periodical precipitation (= primary or ultimate ecofactor), cause seasonal inundations in riverine habitats (e.g., Central Amazonia) or dryness of the soil due to a decreasing river level (e.g., East Zaire). Both situations cause an interruption of the reproduction in carabid beetles during these unfavorable conditions. They induce a gonad dormancy which is regulated by changing temperatures in the habitat (= secondary or proximate ecofactor). For this reason, Central Amazonian inundation forests have been postulated to represent one of the locations at which seasonality has been generated in carabids and other arthropods (Adis, 1992; Erwin, 1998).

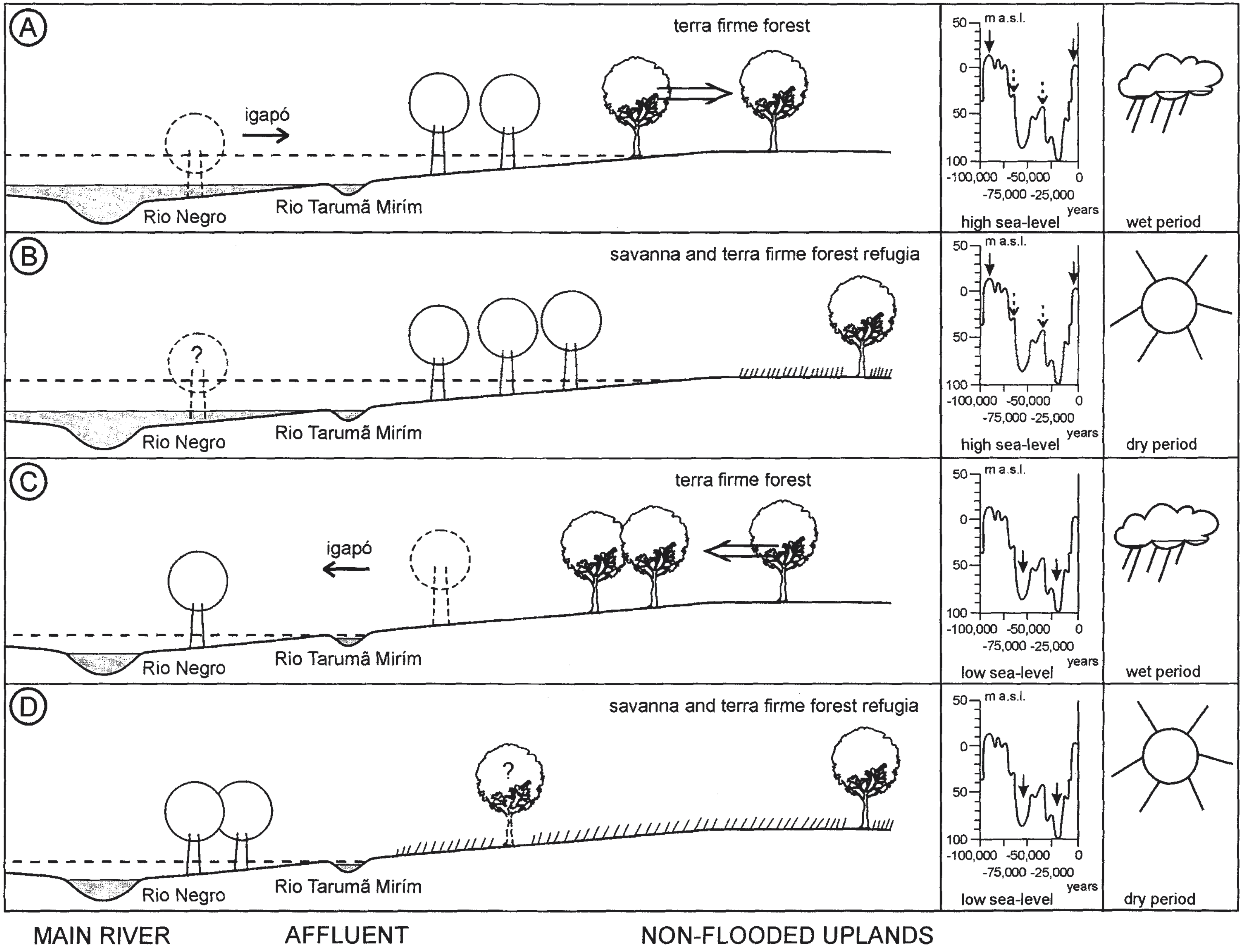
III. SPECIES RICHNESS: CAUSES AND MAINTENANCE

Biological diversification in the Neotropics is the subject of ongoing and frequently antagonistic discussions. Several theories have been proposed to explain the high species richness of Amazonian plants and animals, mostly in a deductive sense, based on selected groups. Studies on distribution patterns aim to detect centers of endemism. It is hoped that biogeographical maps, now also based on cladistic methods, will elucidate evolutionary patterns among these centers. However, it is accepted that data bases are still insufficient (with the possible exception of those for birds) for the type of detailed study needed to understand distribution patterns for most Neotropical groups. Full knowledge of distribution will not be obtained until we understand what sets the limits, for example, the edges of species distribution (Vanzolini and Heyer, 1988).

Past assessments of this type have been based mostly on the presence of Pleistocene refugia (see Section II), also called the "Haffer effect." This is the theory that Neotropical forest fragmentation occurred during northern ice advances due to Milankovitch cycles and the temporary existence of forests and grasslands (savanna), called "island" refugia, that caused temporal centers of endemism (Haffer as cited in Prance, 1982; Haffer, 1996).

Studies on 162 species of Neotropical butterflies (Nymphalidae) by Brown indicated the existence of 50 principal forest centers of evolution and endemism. Smaller areas, recognizable as 38 forest refuges, presumably acted during the last major dry, cold spell in the Quartenary (20,000–13,000 years B.P.). Studies by Erwin and Pogue on arboricolous carabid beetles, adapted to life in Neotropical forest canopies (Genus *Agra*), showed that species group distribution is much broader. In this case, areas considered to be centers of endemism are geographically larger than that in butterflies and vertebrates (Prance, 1982; Vanzolini and Heyer, 1988).

Haffer's forest refuge model as stated deals only with non-flooded terra firme forests, their fragmentation, and/or replacement by savannas (*sensu lato*). There has been no indication that riverine forests disappeared along the lengths of rivers, but rather it appears that savannas were behind such forests as they are today. For this reason, Erwin and Adis consider Central Amazonian floodplain forests, especially along black-water rivers in the Negro delta (called igapó), to be short-term refuges and long-term evolutionary centers (Prance, 1982; Adis, 1984). Flood cycles greatly influence speciation rate and dispersion of arthropods. Evidence from Rio Tarumã Mirím, an affluent of the Rio Negro above Manaus, suggests alternation of two types of inundation: annual flooding and continental water table rise during the past million years. The area examined originally was covered by a non-flooded terra firme forest. Periods of high sea level caused a back-up of main riverine courses in Amazonia. During these periods, the affluent Tarumã Mirím was high enough that annual water fluctuations of the Rio Negro led to an inundation lasting several months in the forest under study. The first inundation was several million years ago. Since that time, various periods of high sea level (according to Fairbridge) have occurred (e.g., 250,000, 170,000, and 85,000 years ago; Fig. 1), with annual flooding of 10,000–30,000 years duration. Approximately 6000 years ago, the sea level reached a height of about 5 m below mean sea level. Since then, the examined forest has been inundated again. Flood cycles were responsible for extant vegetation formation in the igapó. During high sea level, the (lower) bank of the Rio Negro suffered longer periods of inundation and, according to Sioli, possible formations of large "ria-lakes." Igapó forests were drowned (at least partially) and backed up minor tributaries which subsequently had annual flooding (e.g., Rio Tarumã Mirím; Fig. 1A). They replaced non-flooded terra firme forests that previously occupied the area. Intermittent dry spells during high sea level periods may have caused formation of savanna and non-flooded terra firme forest refugia (Fig. 1B), as postulated by Haffer. During low sea level periods, annual inundation occurred only along main rivers (e.g., the Rio Negro); this has been the case for at least 1 million years, probably including all glacial advances of the Pleistocene and perhaps even since the Andes took their present configuration. Non-flooded vegetation on terra firme ("uplands") spread extensively, forcing igapó forests along affluents to retreat to main rivers


A
terra firme forest
igapó
Rio Negro
Rio Tarumã Mirím
m a.s.l.
50
0
50
100
-100,000 -50,000 0
-75,000 -25,000
years
high sea-level
wet period
B
savanna and terra firme forest refugia
?
Rio Negro
Rio Tarumã Mirím
high sea-level
dry period
C
terra firme forest
igapó
Rio Negro
Rio Tarumã Mirím
low sea-level
wet period
D
savanna and terra firme forest refugia
?
Rio Negro
Rio Tarumã Mirím
low sea-level
dry period
MAIN RIVER
AFFLUENT
NON-FLOODED UPLANDS

(Fig. 1C). Intermittent dry periods, according to Haffer, again caused savanna formation with possible non-flooded forest refugia on terra firme and "igapó forest refuges" (Fig. 1D).

Flood cycles have a dramatic impact on inundation forest faunas. This is especially important when two kinds of flooding patterns manifest themselves through time. In the greater region of Amazonia, there could have been periods favoring lakeshore species, then inundation forest species, and then riparian species (for discussion of carabid beetles, see Erwin and Adis as cited in Prance, 1982). Large waterbodies and drying affluents, with possibly additional regression of non-flooded forests on terra firme in both cases, cause interruption and isolation of igapó forests. Species vagility then seems to become a crucial factor in sorting out and isolating gene pools. Thus, small barriers may be highly disruptive to populations of individuals with low vagility, particularly non-flying soil arthropods. Supposing that intense competition and predation promote rapid speciation (especially in igapó forests, in which flooding mixes terrestrial and canopy biotas during half the year) and that the isolation of igapó forests is sufficiently long for rapidly evolving species, this model represents another key to Amazonian species richness. In fact, recent data indicate that (i) igapó forests retain many endemic species; (ii) the fauna of igapó forests is different in composition and activity patterns when compared to that of adjacent non-flooded terra firme forests; and (iii) in igapó forests across the Rio Negro and its tributaries, taxonomically different subspecies and species are found.

A third theory considers the high plant and animal diversity of Amazonian non-flooded terra firme to be a response to extremely low nutrient concentrations in geochemically impoverished ecosystems (Fittkau, 1973; Klinge, 1973). Species richness and simultaneously occurring high diversity of moist tropical ecosystems do not reflect high nutrient supply but rather a mode of adaptation to continuous restriction of nutrients or food substances under otherwise permanently favorable living conditions. Plants and animals are assumed to act as highly efficient "nutrient traps." During evolution of multiple life forms, shortages must have been of significant importance at all times.

A model group supporting this theory are the carabid communities on fig fruit falls. Paarmann and Adis sampled 8962 beetles on 65 fruit falls from 10 fig species between 1991 and 1996 in a non-flooded terra firme forest (Reserva Ducke) near Manaus. Eight of the 36 species collected represented the spermatophagous genus *Notiobia*, which accounted for 92% of all carabid beetles obtained. Only one *Notiobia* species was dominant. The abundance distribution of the eight *Notiobia* species was very similar on fruit falls of the two most common fig species as well as on fruit falls of the remaining fig species. The dominance structure varied considerably between individual fruit falls. Also, the abundance distribution changed during the course of a single fruit fall. Only 2 of the 8 *Notiobia* species were found to be specialized fig seed feeders, able to reproduce only on fig fruit falls (Vanicek *et al.*, 1994). The remaining 6 species of this genus use fig fruit falls as stepping stones between fruit falls of their host trees, which are separated by time and space.

Another theory for the high within-community diversity of tropical arthropods is that continuous stochastic local disturbances in nature are assumed to prevent the achievement of any long-term equilibrium (climax) state. These so-called stochastic non-equilibrium models assume that the presence of a species at a vacant site is important. This may represent an advantage against all species that arrive later. Niche overlaps are assumed to be very common within species-rich communities. As a result, neither successional nor a climax community can emerge. This is in contrast to the so-called deterministic equilibrium models, which are based on the ecological niche. Each organism maintains a defined position in its environment and, driven by competition, the system goes through defined successional stages and a structurally predictable climax equilibrium results (Linsenmair, 1990).

Examples which clearly support one or the other mechanism are not available for Amazonian arthropods. For example, concerning the diversity of the seed-feeding ground beetle community of the genus *Notiobia*, they are specialized on certain seeds as food for a successful larval development (deterministic process) and yet the non-fig seed specialists consume fig seeds to survive periods of food shortage. The necessity to do so limits the tendency to specialize. If stochastic processes do play a role in maintaining

FIGURE 1 Postulated formations and shifts of black-water inundation forests along main rivers and affluents of Central Amazonia within the past 100,000 years due to changes in sea level and climate (see text for further explanation; modified with permission from Adis, 1984 with kind permission from Kluwer Academic Publishers).

the diversity of the seed-feeding *Notiobia* species, this role is probably less important. Thus, they are "chance specialists" adapted to the unpredictability of fruit falls. The fig seed specialists belong to the "touring group," moving from one fig fruit fall to the next, depending on a sufficient density and an annually equal distribution. The aseasonal fruiting of figs is caused by their pollination biology.

IV. BIOTOPE AND HABITAT SPECIFICITY OF SPECIES

Conservation or sustainable use of Amazonian ecosystems demand information on biodiversity. Due to the lack of basic comparative data on arthropods, recent recommendations mostly relate to the high diversity and rarity of species found in certain regions, which are called areas of "high biological value" (Anonymous, 1991; Dinerstein *et al.*, 1995; Kress *et al.*, 1998), rather than to specific indicator species of biotopes and/or habitats revealing endemism and possible centers of radiation. Two long-term studies on pseudoscorpions and tiger beetles demonstrate the validity for such information.

Pseudoscorpions were studied intensively in Central Amazonian biotopes between 1975 and 1990. The 35,000 specimens collected represented 26 genera and 60 species (Adis and Mahnert, 1990). Of these, 29 species (48%) occurred exclusively in non-flooded terra firme forests and 25 species (42%) were restricted to floodplain forests. Species from terra firme forests were more terricolous, with 69% living in litter and soil. Species from floodplain forests were predominantly arboricolous, with 88% living in the trunk and/or canopy region. Differences in habitat selection are attributed to the impact of the flood pulse in Amazonian floodplains, i.e., long-term inundation of 5–7 months' duration. Only 11 species were found in forests inundated by white-water (Rio Solimões-Amazonas) compared to 20 species found in black-water forests (Rio Negro). This is attributed to the differences observed between biotopes (flora and soil structure) as well as between the two river systems (current and sediment load; Adis as cited in Junk, 1997).

Data-bases such as these help to answer general questions about specific biotopes. For example, the terrico-

TABLE II

Species of Tiger Beetles (Carabidae: Cicindelinae) Sampled in White-Water Floodplains (Várzea) of the Rio Solimões-Amazonas and on Non-flooded Terra Firme Near Manaus, Central Amazonia[a]

	Várzea floodplains		Terra firme uplands
Species	9		15
Forests	3		7
Joint		1	
Open areas	6	0	8
Joint			
	Cylindera (Plectographa) suturalis (Fabricius) (O, d) *Megacephala (Tetracha) sobrina punctata* Castelnau (O, n) *Megacephala (Tetracha) spinosa* (Brullé (O, n) *Megacephala (Phaeoxantha)* aequinoctialis Dejean (O, n) *Megacephala (Phaeoxantha) klugi* Chaudoir (O, n) *Odontocheila confusa* (Dejean) (F, d) *Pentacomia (Mesacanthina) cribrata* (Brullé) (O, d) *Pentacomia (Pentacomia) egregia* (Chaudoir) (F, d) **Pentacomia (Poecilochila) lacordairei** Gory (F, d)		*Aniara sepulcralis* (Fabricius) (O, d) *Brasiella (Cicindela) argentata* (Fabricius) (O, d) *Brasiella (Cicindela) pretiosa* (Dokhtouroff) (O, d) *Cenothyla varians* (Gory) (F, d) *Ctenostoma (Stenoctenostoma) asperulum* Bates (F, ?) *Ctenostoma formicarium* (Fabricius) (F, ?) *Cylindera (Cylindera) morio* (Klug) (O, d) *Megacephala (Tetracha) bilunata* Klug (O, n) *Odontocheila cayennensis* (Fabricius) (F, d) *Odontocheila chrysis* (Fabricius) (O, d) *Odontocheila luridipes* (Dejean) (F, d) *Odontocheila margineguttata* (Dejean) (O, d) *Odontocheila nigrotarsalis* Horn (F, d) **Pentacomia (Poecilochila) lacordairei** Gory (F, d) *Pentacomia (Poecilochila) ventralis* (Dejean) (O, d)

Note. Abbreviations used: F, forests; O, open areas; d, diurnal; n, nocturnal.

[a] Modified from Adis *et al.* (1998b).

lous pseudoscorpions of terra firme forests on white-sand soil in Central Amazonia (called campinarana) represent the species spectrum of primary terra firme forests on yellow latosol and show no endemic species (Adis and Mahnert, 1993). This reinforces the geological results that indicate that campinarana does not represent forests along former river-beds which have dried up, as previously suggested, but rather the final stage of podzolization (i.e., the transformation of clayey latosols to white-sand podzols by long-term weathering and leaching processes).

Of the 24 species of tiger beetles (Carabidae: Cicindelinae) studied between 1991 and 1998 in the Manaus area (Table II), 15 were found in non-flooded terra firme areas. Of these, 7 species lived in forests. Five of them, being diurnal, inhabited the floor (Fig. 2: CVA, OCA, OLU, ONI, and PLA), and 2 species were canopy dwellers (CAS and CFO). Eight species, 1 nocturnal (MBI) and 7 diurnal (ASE, BAR, BPR, CMO, OCH, OMA, and PVE), lived in open areas. Three of these were found on mostly bare latosol (BAR, OCH, and PVE). The other 5 species occurred on mostly bare white sand (ASE, BPR, CAR, CMO, and MBI). The species assemblage in open areas indicated an impact by human intervention: *Pentacomia ventralis* (PVE) is known as a species that inhabits extensive cleared forest sites and *Odontocheila margineguttata* (OMA) occurs in secondary forests. Another 9 tiger beetle species were recorded from white-water floodplains (várzea; Table II). Of these, 3 diurnal species inhabited inundation forests (Fig. 3: OC, PE, and PL). Six species, 2 diurnal (CS and PC) and 4 nocturnal (MA, MK, MP, and MS), lived in open areas. Four were found on bare beaches (CS, PC, MP, and MS). The other 2 species (MA and MK) inhabited grass-grown areas behind the beaches which were mainly composed of annual and perannual aquatic grasses. Only 1 species, *Pentacomia lacordairei*, occurred in the forests of both non-flooded terra firme and floodplains (Table II).

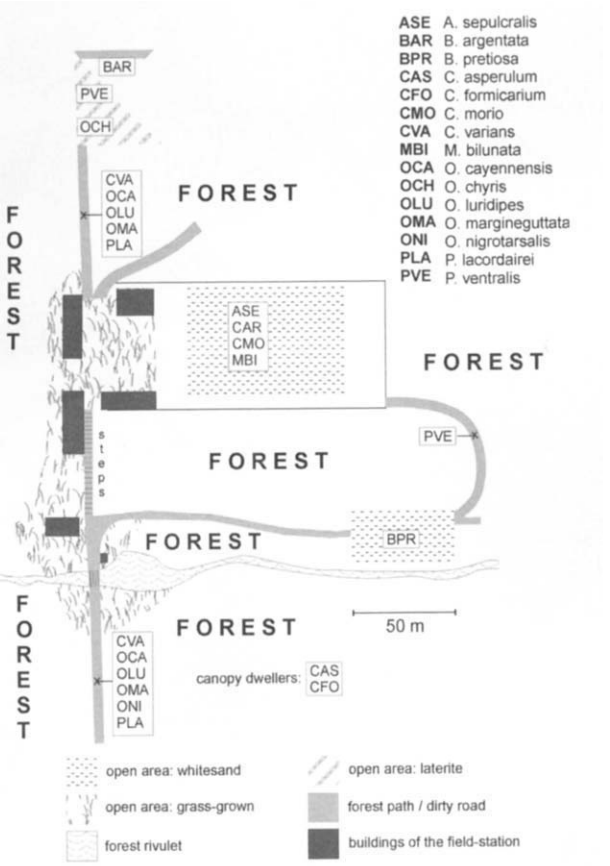


FIGURE 2 Species of tiger beetles (Carabidae: Cicindelinae) occurring in the vicinity of the field station at the Adolpho Ducke forest reserve (Reserva Ducke) near Manaus, Central Amazonia (reproduced with permission from Adis *et al.*, 1998b).

Data clearly show the biotope and habitat specificity of these tiger beetle assemblages. Moreover, the life-cycle of species on non-flooded terra firme differed greatly from those inhabiting floodplains, which indicates a high adaptation to the respective biotope and habitat. Data such as these might also help to elucidate questions of general interest. For example, the dichotomy of forest or open-habitat assemblages might be important for understanding tiger beetle distribution in Amazonia. Similar patterns were reported by Heyer for frogs east of the Andes (Vanzolini and Heyer, 1988). Data also support the potential use of tiger beetles as bioindicators for monitoring the degradation and regeneration of Amazonian forests. In Venezuela, forest-floor species assemblages changed significantly with the degree of forest disturbance, and each stage of disturbance was characterized by a particular subset of species (Rodríguez *et al.*, 1998).

V. MORPHOLOGICAL SPECIES

In millipedes, taxonomy is mainly based on the structure of genitalia in males, the gonopods. However, in the process of adaptation to new environments, "pioneers" apparently change their external structure more quickly than their gonopods, which results in difficulties in defining species. According to Hoffman (1990), the diversification of the external body form with little modification of gonopod structures in the order Polydesmida seems to be associated with those taxa (usually genera) which seem to have recently occupied a new area or biotope independently of close relatives and are indulg-

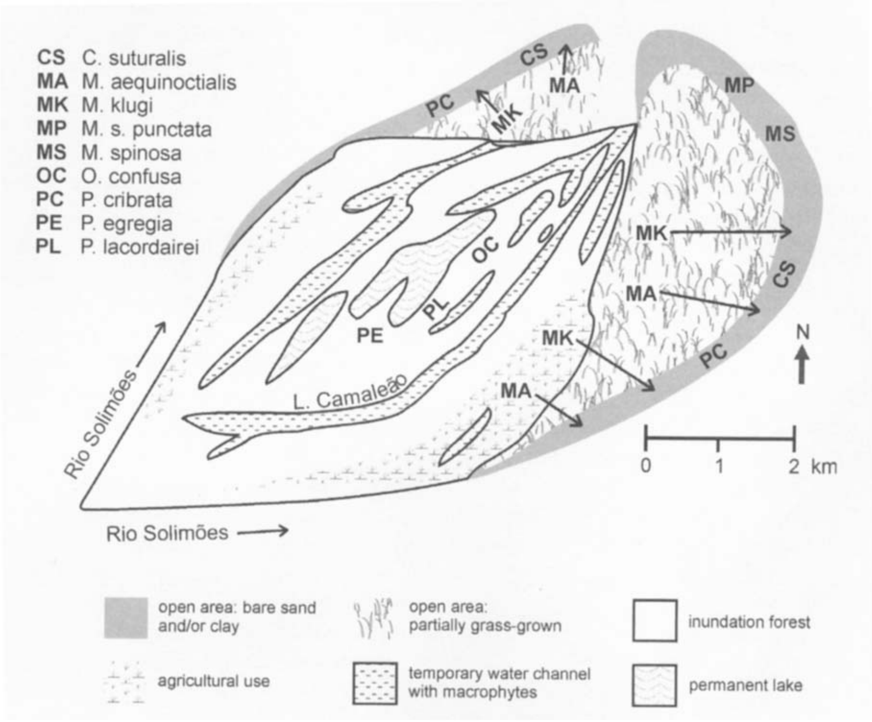


FIGURE 3 Species of tiger beetles (Carabidae: Cicindelinae) found on Ilha de Marchantaria, the first island in the Rio Solimões near Manaus, Central Amazonia (reproduced with permission from Adis *et al.*, 1998b).

ing in a burst of adaptive radiation unimpeded by sibling competition. It is presumed that after an initial period of great diversification some selection of successful lineages would occur. During this time, body form types, stabilized by adaptational factors, would tend to remain constant. The generic variability, however, might be expressed much more rapidly in structures such as genitalia, which are less directly influenced by environmental constraints. Thus, a period would ensue in which these species groups (or genera) would consist of species readily distinguishable by highly distinctive gonopods. Such a stasis would endure for a long time unless it were again possible for a fragment of a group (probably migrant populations of one or two species) to escape into a new area or new ecological niche and begin the cycle anew. This situation seems to be different from the punctual evolution of new clades in that two basically different character systems alternate in phase with the chronological status of the organisms in "new" and "old" territory and thus underscore the principle that phylogeny cannot be adequately interpreted outside the context of biogeography.

New insights derived from studies on the Neotropical genus *Pycnotropis* support these postulations. Among the 26 species known, only 3 are restricted to Amazonian white- and mixed-water inundation forests, whereas the remainder are found in non-flooded terra firme forests. Biogeographically, the history of *Pycnotropis* can be viewed as one implying the origins in the western Andes, the postulated center of origin (and radiation) of the entire subfamily Amplininae, with subsequent waves of downstream dispersal throughout the Amazon Basin. The present-day distribution of *Pycnotropis* is the result of numerous, apparently ongoing vicariance events involving repeated gene flows downstream of the Amazon and its tributaries (radiation corridors) from source areas, with colonization and recolonization of inundation forests and/or non-flooded terra firme habitats. The genus is still in a stage of very active speciation because at least some of its species display pronounced variation in both external and gonopod characters. This is supported by preliminary tests concerning genetic variation in *Pycnotropis tida*, a species that has been described twice due to variation in morphological characters. The pattern of a somewhat "unsettled" speciation process seems to contrast the one observed in the chelodesmid genus *Camptomorpha*, which likewise is species rich in the Andes and abundantly represented in the Amazon Basin. It is, however, a good example of a taxon in the "stable body–variable gonopod" stage, with each species being distinct morphologically (Golovatch *et al.*, 1998).

VI. GENETICAL SPECIES

Comparative phenological studies on arthropod species which inhabit both floodplain and non-flooded terra firme forests in Central Amazonia suggested an uni- or bivoltine life cycle of populations in annually inundated forests and a plurivoltine mode of life of populations in terra firme forests. Representatives of both populations were morphologically alike; hence, they represented the same species (e.g., in Pseudoscorpiones, Diplopoda, Symphyla, and Archaeognatha).

In the case of *Neomachilellus scandens*, a member of the Meinertellidae (Archaeognatha, Insecta), the two populations in terra firme and floodplain forests were first postulated to be "biotope-specific races." Morphological characters that would justify separation into two species could not be found by the taxonomist, at least with traditional research methods. However, protein analyses by means of electrophoresis (testing 15 enzymes) revealed that there is no gene flow between populations from floodplain and terra firme forest types, even when they are only 50 m distant from each other. Thus, this "species" was proposed to represent two different species based on eco-phenological and genetic characteristics. Without this genetical background, problems arise if the two genetical species occur sympatrically because, due to the intermixture of data, the eco-phenological differences are hardly or not at all recognizable. This was the case in a mixed-water inundation forest near Manaus (Wolf and Adis, 1992). Until now, culturing and breeding of Amazonian Meinertellidae was unsuccessful; thus, interbreeding experiments could not be performed.

The polydesmidan millipede *Pycnotropis tida* (see Section V) represents a second example. In the Manaus area, populations which inhabit secondary terra firme forests breed throughout the year, whereas populations of white- and mixed-water inundation forests show a defined, univoltine reproduction during the non-aquatic phase (Vohland and Adis, 1999). The genetic data were obtained from allozyme analyses (testing 14 enzymes by electrophoresis) and from a specific satellite DNA. In this case, results suggested that individuals from the two biotopes represented populations of the same species, although genotypic structures among and within local populations indicated processes of ongoing genetic differentiation (Bachmann *et al.*, 1998). A successful interbreeding thereafter reinforced the genetical results.

For taxonomists, the use of genetic techniques and the consideration of ecological as well as ethological data represent the final, although mandatory, stage of analysis in systematic biology. In the case of Amazonian Pseudoscorpiones and Symphyla, both genetic and breeding exercises have not been successful.

VII. PROSPECTS

Amazonian terrestrial arthropods are special: They have adapted in some way to their environment. Only detailed studies at the species level will reveal the "strategies" which have evolved. For example, there are (i) grasshoppers on aquatic macrophytes in the Rio Amazonas with spiny front legs, like mantids, that follow their prey into tree canopies during high water (*Phlugis teres*, Tettigoniidae); (ii) a small flood-resistant "terrestrial" millipede which lives up to 11 months submerged in inundation forests along the Rio Negro, breathing under water by means of a plastron and feeding on algae (*Gonographis adisi*, Pyrgodesmidae); and (iii) larvae of a carabid beetle (normally a predaceous taxon) which accomplish their development feeding on fig seeds (*Notiobia flavicintus*) but apparently depend on larvae of their sister species (*N. pseudolimbipennis*) which are able to open the seed shells with their morphologically larger mandibles. Some insights into this fascinating world have been achieved since Bates voyages on the "River Amazons." However, it took about 150 years to adopt his suggestion to study 1 acre of rain forest in more detail. Two approaches are currently being taken. One is to develop an All-Taxa-Biodiversity-Inventory and the other to develop an All-Biota-Taxon-Inventory (Platnick, 1999). Both enterprises require an intensified training of tropical taxonomists and eco-entomologists.

Acknowledgments

I gratefully acknowledge the valuable comments received on a draft of this article by Terry L. Erwin, Sergei I. Golovatch, Richard

L. Hoffman, Wilfried Paarmann, Norman I. Platnick, and Harald Sioli.

See Also the Following Articles

AMAZON ECOSYSTEMS • ARACHNIDS • BEETLES • FOREST CANOPIES, ANIMAL DIVERSITY • HYMENOPTERA • MYRIAPODS

Bibliography

Adis, J. (1984). "Seasonal igapó"-forests of Central Amazonian blackwater rivers and their terrestrial arthropod fauna. In *The Amazon. Limnology and Landscape Ecology of a Mighty Tropical River and Its Basin. Monographiae Biologicae* (H. Sioli, Ed.), pp. 245–268. Junk, Dordrecht.

Adis, J. (1992). Überlebensstrategien terrestrischer Invertebraten in Überschwemmungswäldern Zentralamazoniens. *Verh. Naturwiss. Ver. Hamburg (NF)* 33, 21–114.

Adis, J., and Harvey, M. S. (2000). How many Arachnida and Myriapoda are there worldwide and in Amazonia. *Stud. Neotrop. Fauna Environ.*, in press.

Adis, J., and Mahnert, V. (1990). On the species composition of Pseudoscorpiones (Arachnida) from Amazonian dryland and inundation forests in Brazil. *Rev. Suisse Zool.* 97, 49–53.

Adis, J., and Mahnert, V. (1993). Vertical distribution and abundance of pseudoscorpions (Archachnida) in the soil of two different Neotropical primary forests during the dry and rainy seasons. *Mem. Qd Mus.* 33(2), 431–440.

Adis, J., Harada, A. Y., da Fonseca, C. R. V., Paarmann, W., and Rafael, J. A. (1998a). Arthropods obtained from the Amazonian tree species "Cupiuba" (*Goupia glabra*) by repeated canopy fogging with natural pyrethrum. *Acta Amazonica* 28(3), 273–283.

Adis, J., Paarmann, W., Amorim, M. A., Arndt, E., and da Fonseca, C. R. V. (1998b). On occurrence, habitat specificity and natural history of adult tiger beetles (Coleoptera: Carabidae: Cicindelinae) near Manaus, Central Amazonia, and key to the larvae of tiger beetle genera. *Acta Amazonica* 28(3), 331–344.

Anonymous (1991). Workshop 90: Biological Priorities for Conservation in Amazonia (map and explanatory text). Conservation International, Washington, DC.

Bachmann, L., Tomiuk, J., Adis, J., and Vohland, K. (1998). Genetic differentiation of the millipede *Pycnotropis epiclysmus* inhabiting seasonally inundated and non-flooded Amazonian forests. *J. Zool. Syst. Evol. Res.* 36, 65–75.

Dinerstein, E., Olson, D. M., Graham, D. J., Webster, A. L., Primm, S. A., Bookbinder, M. P., and Ledec, G. (1995). *A Conservation Assessment of the Terrestrial Ecoregions of Latin America and the Caribbean.* The World Bank, Washington, DC.

Erwin, T. L. (1991). An evolutionary basis for conservation strategies. *Science* 253, 750–752.

Erwin, T. L. (1998). Evolution at the equator: Arboreal and alticolous beetles and their taxon pulses with descriptions of a new *Agra* subclade and its species (Coleoptera: Carabidae: Lebiini). *Mus. Reg. Sci. Nat. Torino,* 491–510.

Fittkau, E. J. (1973). Artenmannigfaltigkeit amazonischer Lebensräume aus ökologischer Sicht. *Amazoniana* 4(3), 321–340.

Golovatch, S. I., Vohland, K., Hoffman, R. L., Adis, J., Mármol, A., Bachmann, L., and Tomiuk, J. (1997). Review of the Neotropical millipede genus *Pycnotropis* Carl, 1914 (Diplopoda, Polydesmida, Aphelidesmidae). *Amazoniana* 15(1/2), 67–102.

Haffer, J. (1996). Time's cycle and time's arrow in the history of Amazonia. In *Uma Estratégia Latino-Americana para a Amazônia. Vol. 1.*, (C. Pavan and M. C. de Araujo, Eds.), pp. 21–49. Min. do Meio Amb., dos Rec. Hídricos e da Amaz. Legal., Brasilia.

Hoffman, R. L. (1990). Myriapoda 4. Polydesmida: Oxydesmidae. In *Das Tierreich* (H. von Wermuth and M. Fischer, Eds.), Vol. 107. de Gruyter, Berlin.

Junk, W. J. (Ed.) (1997). The Central Amazon floodplain. Ecology of a pulsing system. *Ecological Studies, Vol. 126.* Springer, Berlin.

Klinge, H. (1973). Struktur und Artenreichtum des zentralamazonischen Regenwaldes. *Amazoniana* 4(3), 283–292.

Kress, W. J., Heyer, R., Acevedo, P., Coddington, J., Cole, D., Erwin, T. L., Meggers, B. J., Pogue, M., Thorington, R. W., Vari, R. P., Weitzmann, M. J., and Weitzmann, S. H. (1998). Amazonian biodiversity: Assessing conservation priorities with taxonomic data. *Biodiv. Conserv.* 7, 1577–1587.

Linsenmair, K. E. (1990). Tropische Biodiversität: Befunde und offene Probleme. *Verh. Dtsch. Zool. Ges.* 83, 245–261.

Lourenço, W. (1994). Biogeographic patterns of tropical South American scorpions. *Stud. Neotrop. Fauna Environ.* 29(4), 219–231.

Penny, N. D., and Arias, J. R. (1982). *Insects of an Amazon Forest.* Columbia Univ. Press, New York.

Pereira, L. A., Foddai, D., and Minelli, A. (1997). Zoogeographical aspects of Neotropical Geophilomorpha (Chilopoda). *Ent. Scand. Suppl.* 51, 77–86.

Platnick, N. I. (1999). *Dimensions of biodiversity: Targeting megadiverse groups.* In *The Living Planet in Crisis: Biodiversity Science and Policy* (J. Cracraft and F. Grifo, Eds.), pp. 33–52. Columbia Univ. Press, New York.

Prance, G. T. (Ed.) (1982). Biological diversification in the tropics. In *Proceedings of the Fifth International Symposium of the Association for Tropical Biology.* Columbia Univ. Press, New York.

Rodríguez, J. P., Pearson, D. L., and Barrera, R. (1998). A test for the adequacy of bioindicator taxa: Are tiger beetles (Coleoptera: Cicindelidae) appropriate indicators for monitoring the degradation of tropical forests in Venezuela? *Biol. Conserv.* 83, 69–76.

Sturm, H. (1984). Zur Systematik, Biogeographie und Evolution der südamerikanischen Meinertellidae (Machiloidea, Archaeognatha, Insecta). *Z. Zool. Syst. Evol.* 22, 27–44.

Vanicek, M., Adis, J., and Paarmann, W. (1994). Untersuchungen zur Biologie dreier Laufkäferarten (Coleoptera, Carabidae, Harpalini) aus amazonischen Regenwäldern. *Andrias* 13, 161–168.

Vanzolini, P. E., and Heyer, W. R. (Eds.) (1988). *Proceedings of a Workshop on Neotropical Distribution Patterns.* Academia Brasileira de Ciências, Rio de Janeiro.

Vohland, K., and Adis, J. (1999). Life history of *Pycnotropis tida* (Diplopoda: Polydesmida: Aphelidesmidae) from seasonally inundated forests in Amazonia (Brazil and Peru). *Pedobiology* 43, 231–244.

Wolf, H. G., and Adis, J. (1992). Genetic differentiation between populations of *Neomachilellus scandens* (Meinertellidae, Archaeognatha, Insecta) inhabiting neighbouring forests in Central Amazonia. *Verh. Naturwiss. Ver. Hamburg (NF)* 33, 5–13.

ASIA, ECOSYSTEMS OF

Elgene O. Box and Kazue Fujiwara
University of Georgia and Yokohama National University

GLOSSARY

angiosperms Flowering vascular plants in which the seeds, which are produced from the ovules of the flowers, are enclosed in fruits developed from the matured ovaries; the largest class of vascular plants, as opposed to cryptogams (mosses, etc., which do not produce seeds) and conifers (in which the seeds are not enclosed in fruits).

biodiversity Biological diversity, i.e., the variety of living organisms in an area, including the variety in genes, species, functional types of organisms, and ecosystems.

biome A major terrestrial ecological community and landscape type, characterized by more or less uniform physiognomy of its potential natural vegetation and with characteristic fauna and flora, such as the tropical rainforest, the warm deserts, or the temperate grasslands.

boreal Pertaining to the northern high latitudes (but not polar), regions which contain large continental landmasses and thus continental climates with moderate summers but long, severely cold winters (from Greek *boreas,* the north wind).

ecosystem A community of organisms (plants, animals, and microbes) and their physical environment interacting as an ecological unit, such as a lake, a wetland, a forest, or an agricultural landscape.

endemic Occurring only within a comparatively restricted geographic range within a specific region.

global circulation The regular global circulation pattern of the earth's atmosphere, which generates the world's basic climate types; the system involves an Intertropical Convergence zone of low pressure near the equator, subtropical high-pressure belts near the Tropics of Cancer and Capricorn, trade winds flowing from these high-pressure belts toward the equatorial low, and westerly winds in the midlatitudes.

monsoon system A wind system covering the eastern half of Asia in which winter cooling of the large landmass produces strong, stable high pressure, with clear skies and outward flow of cold, dry air, and summer warming produces low pressure, drawing wet air masses, with clouds and rain, inward from the adjacent oceans.

Pleistocene glaciation The expansion of ice sheets over the large northern continents, especially northern

Encyclopedia of Biodiversity, Volume 1

Europe and North America (excluding the northwest) during the "Ice Ages."

potential natural vegetation The vegetation cover which would develop naturally in an area and become stable (not replaced by a subsequent stage) if all outside disturbances were eliminated.

species richness The total number of species of an area or ecosystem.

temperate Pertaining to the climates and landscapes of the midlatitude regions, which are seasonally warmer and cooler, with winter frost even in most coastal areas and at least partial dormancy or collapse of the vegetation.

tropical Astronomically, the region lying between the Tropic of Cancer (23½°N) and Tropic of Capricorn (23½°S); more generally, involving the climates and landscapes characteristic of this region, which are essentially frost free in the lowlands and permit biological activity as long as water is available.

vegetation The total plant cover of an area.

zonation The tendency of climate, soil, and natural landscape types to occur in distinct latitudinal zones (e.g., tropical, subtropical, and temperate), as generated by the global atmospheric circulation system.

ASIA is the world's largest continent, extending from the Mediterranean Sea in the west to the Japanese Archipelago in the east, and from polar landscapes in northern Siberia to hot deserts in the southwest and tropical regions south of the Himalaya and China. Tropical Asia includes perhumid equatorial rainforest climates as well as seasonally wet and dry tropical savannas, raingreen woodlands, and monsoon forests, all extending to some degree from India to the Philippines and East Indies. Dry regions of Asia include an eastward extension of the hot, subtropical Sahara desert into southwestern Asia, large expanses of higher latitude interior deserts with cold winters, and also a small "mediterranean" region and an east–west strip of temperate grassland (steppe) which spans almost the entire continent. The main area of temperate forests is in East Asia, with deciduous and mixed forests in the north, rich evergreen "laurel" forests in the south, and extensive areas of mostly secondary pine forests throughout. Mountain areas of Asia have many endemic species, especially relict conifers, and the mountains of East Asia represent the only truly large mountainous area in the warm-temperate and humid subtropical zones of either hemisphere. Northern Asia was less glaciated than Europe or North America and contains enormous expanses of relatively diverse Siberian conifer forests, polar and upland tundra, and Lake Baikal (the world's deepest). Wetlands in Asia include mangroves, coastal strand forests, salt marshes and estuaries, and terrestrial swamp forests, marshes, and bogs, most of which are highly productive and represent critical habitat for terrestrial as well as aquatic animals. Despite long histories of human habitation and landscape alteration, Asia retains the highest biodiversity of any continent, due not only to its size but also to its climatic and topographic complexity and its complex geological and evolutionary history. This biodiversity is threatened in Asia as elsewhere by human overpopulation and overdevelopment.

I. ASIA THE REGION

Asia is traditionally separated from Europe, with which it shares the world's single largest land-mass Eurasia (Fig. 1). Even by this cultural definition, however, Asia remains the world's largest continent and extends almost halfway around the world, from a small western coastline on the Mediterranean Sea (about 25°E in Turkey) eastward to the Bering Strait (170°W) separating it from Alaska. Asia also spans almost the entire latitudinal range of the Northern Hemisphere, from a northern coastline in the Arctic zone to islands lying slightly south of the equator in Indonesia. Asian superlatives include the world's

- highest mountain (Mt. Everest, 8848 m) and mountain range (Himalaya);
- largest highland plateau (Tibet);
- lowest terrestrial elevation (Dead Sea, −400 m);
- largest (Caspian Sea, 143,200 km^2) and deepest (Lake Baikal, 1620 m) lakes; and
- coldest, most continental region except for Antarctica (northeastern Siberia), with mean monthly temperature ranging from 20 to −60°C and extremes to −80°C.

Also found in Asia are

- the world's third (Borneo) and fifth (Sumatra) largest islands;
- four of the world's seven rivers over 5000 km long (Yangtze, Ob'-Irtysh, Yenisey-Angara, and Huang-He);
- one of the world's three regions with essentially no rainfall (Tarim Basin); and
- two of the world's three stations with more than 10 m of average annual rainfall (windward slopes

FIGURE 1 Location and main physiographic features of Asia. Asia extends almost halfway around the world, from Turkey and Arabia in the west to the Bering Strait in the east. It also extends from the Arctic Ocean in the north to slightly south of the equator in Indonesia. Middle Asia extends across the Turanian Basin from the Caspian Sea to the Tien-Shan and Altai mountains and is largely steppe and desert. Central Asia includes the high, cold Tibetan Plateau and the desert basins to its north, on into the steppes of Mongolia. Siberia (Asiatic Russia) is the mostly forested region north of the east–west mountains and Middle Asian deserts. Other regions and the geological development of Asia are described in the text.

of easternmost India at Cherrapunji and Yakushima Island of Japan).

In addition, Asia has been home to some of the oldest human cultures and to areas of very long, continuous human habitation, especially in the Middle East, India, and Southeast and East Asia.

Asia represents the largest part of Laurasia, the northern half of the supercontinent Pangaea, which existed in the late Paleozoic and broke up during the Mesozoic (Table I; Raven and Axelrod, 1974). Eurasia began to separate from North America around 180 million years ago, but contact was reestablished in the Cretaceous when Siberia approached Alaska, forming Beringia. At about this time, southwestern Asia also began to be influenced by northward-moving Africa, a portion of Gondwana, the southern part of Pangaea.

The present-day geologic and physiographic struc-

TABLE I

Some Major Events in Asia's Geological and Evolutionary History[a]

		Geological and climatic events	Biotic events
Paleozoic Era (>225 million BP)			
Silurian	425–400 m	Climate similar to present?	First vascular and land plants
Carboniferous	345–280 m	Warm and humid	Forests of ferns, horsetails, etc.
Permian	from 280 m	Pangaea; glaciation at S pole	First ginkgos and mosses
Mesozoic Era (225–65 million BP)			
Triassic	225–185 m BP	Breakup of Pangaea begins; climate cool and humid	Ginkgos and cycads dominant
Jurassic	185–135 m	Breakup of Pangaea: W. Gondwana	Conifers and cycads dominant
Cretaceous	135–65 m	"Greenhouse" climate, high sea level Major separations Gondwana (130–100 m BP) Eurasia and N. America (180 m, initially) Europe and N. America (80 m; not north) Major contacts SW Asia and Africa (100? m) Siberia and Alaska (70 m)	Angiosperms dominant Modern plant orders exist; some modern families and genera exist
Tertiary Period (of Cenozoic Era)			
Paleocene	65 m BP	Africa and Europe separate temporarily	
Eocene	60–38 m	Final separation of Europe and N America; Beringia becomes warmer Himalaya and central Asia plateaus form as India collides with Asia	Most modern families, many genera exist Madro-Tertiary geoflora First primates
Oligocene	38–26 m	Alpine orogeny stresses Middle Asian mountains	
Miocene	26–7 m	Australia approaches SE Asia; islands form Global climates become cooler and drier	
Pliocene	7–2 m	Arabia splits from Africa and contacts Asia, forcing Turkey westward	Most modern plant taxa; herbaceous plants abundant
Quaternary Period (of Cenozoic Era)			
Pleistocene	2–0.01 m BP 18 k 14–12 k	Glacial and interglacial periods; N Europe glaciated, Siberia less Last glacial maximum: sea 150 m lower, land bridges and wider coastal plains Warming and glacial retreat	N–S migration of vegetation zones, with loss of taxa unable to cross barriers Tropical areas drier
Holocene	from 10 k	Continued warming	Temperate and high-latitide soils and biomes redevelop
	8–5 k	Hypsithermal period (warmest)	
	5 k	Sea reaches present level; cooling	Paludification in Siberia

[a] Times are in millions (m) or thousands (k) of years before present (BP). Sources: Pearson (1995), Bridges (1990), and Raven and Axelrod (1974).

ture of Asia, however, is dominated by the Himalayan Mountains and Tibetan Plateau, which arose when India, another piece of former Gondwana, collided with Asia in the Eocene. In addition to the Himalaya, a mountain node was created at its western end, from which other major mountain ranges radiate in various directions. The most important of these is the Tien Shan-Altai system, which extends to the north–northeast into Mongolia, dividing interior temperate Asia into

- a western part, called Middle Asia in Russian literature (essentially Turkestan and the Turanian Basin), influenced far inland by westerly winds from the Mediterranean; and
- an eastern part, called Central Asia (essentially the Tibetan Plateau and northwestern China, plus Mongolia) dominated by the Asian monsoon system.

Other radial ranges from the central node include the Hindu Kush mountains to the southwest, which separate Afghanistan from Pakistan; the Paropamisus and Kopet Dagh further west, which separate Afghanistan and Iran to the south from Turkestan to the north; and the Kunlun and Karakoram ranges, which run east–

west across the Tibetan Plateau. South of Tibet, of course, lies the Deccan Plateau of peninsular India, with its Eastern and higher Western Ghats ranges. Southeast of Tibet, the Yunnan-Guizhou Plateau (southwestern China) represents an upland transition to still mountainous Southeast Asia, with its Dawna and Arakan ranges (Burma) and Annamese cordillera.

In southwestern Asia, the Iranian Plateau (extending into Afghanistan), the Anatolian Plateau of central Turkey, and the Arabian Plateau (peninsula) have all been affected to some extent by the interaction between Africa and Asia. More linear mountain ranges in Asia include the Urals, the Hinggan-Ling and Changbai-Shan in northeastern China, and the Sikhote Alin, Kamchatkan, and other ranges of eastern Siberia. Major interior basins include the Turanian Basin east of the Caspian Sea, the West Siberian Lowland (the world's largest marshland), and the Tarim Basin of Central Asia. Other lowlands include the densely populated East China Plain and many river valleys of Asia: Ganga, Indus, Tigris–Euphrates, Brahmaputra, Irrawaddy, Yangtze, and Huang-He. Less populated lowlands extend along the Siberian coast (with the Taimyr peninsula) and the valleys of the Amur and north-flowing Lena, Yenisey-Angara, and Ob'-Irtysh rivers. The remaining structure of Asia is provided by the volcanic arcs forming the Kuril, Japanese, Okinawan, Philippine, and Indonesian archipelagos, along with continental islands such as Sri Lanka, Hainan, Taiwan, and the smaller Siberian Arctic islands and Andaman-Nicobar chain. All of these areas represent regions of localized environmental conditions and ecosystems, with more or less individual characteristics and diversity.

Climatic zonation in general and major regions in Asia are summarized in Table II. The single most important climatic feature of Asia is the monsoon system, which affects an area from Japan and northeastern China in the east, through China and Southeast Asia, to India and Sri Lanka in the south. The cooling of the vast mid-latitude land-mass in the autumn and winter produces a very large, stable center of high atmospheric pressure over Mongolia and northern China, which feeds cold, dry air outward over the rest of East, Southeast, and South Asia throughout the winter. As a result, and aided perhaps by the many east–west mountain ranges, mean winter temperatures are generally low but extremes are not far below the means. In summer, continental warming causes rising air, which draws wet air masses in off the adjacent Pacific and Indian Oceans, greatly accentuating the rainier season which, due to global atmospheric circulation, would occur in summer over most of this area anyway. The "summer monsoon" arrives as two almost stationary rain fronts which, over the course of about one month, penetrate from Kerala across the rest of India and from Japan across Korea into eastern China. Although the monsoon quickly brings warm rains to south Asia, it brings a cooler, mistier rainy season as it crosses Japan into Korea and China. The cold, dry winter is much more reliable than the summer rains. Heavy winter snow can delay the warming of the landmass to such an extent that the summer monsoon "fails" in places such as India, with disastrous effects on crops.

The Philippines and East Indies are less affected by the monsoon and more by the seasonal shift in global circulation. Near the equator, warm perhumid conditions continue throughout most of the year, with only a short dry season if any at all. The length and degree of the dry season increase away from the equator, reaching several months in the northern Philippines but also in some eastern parts of Indonesia (Lesser Sunda Islands as well as eastern Java and southernmost Borneo and Sulawesi).

Southwestern Asia is not affected by the monsoon but rather by the pervasive effect of the subtropical high-pressure belt. This belt of descending dry air migrates north–south seasonally, creating "Mediterranean" climates with dry summers in Turkey and throughout the Mediterranean borderlands. To the south, however, subtropical high pressure remains the dominant influence throughout the year, causing the desert belt across North Africa, the Arabian Peninsula, and as far east as the Thar desert in Pakistan and western India.

Northern Asia (Siberia) is dominated by its high latitude and large land area, resulting in ultracontinental climates with short summers and long severe winters rivaling the northern polar region in degree of cold.

II. BIODIVERSITY

Angiosperms are thought to have originated in Southeast Asia, and indeed Asia is thought to contain many other foci of more recent biotic radiation as well. During the Pleistocene, when many species were lost from heavily glaciated Europe and North America, glaciation in Asia was less extensive, especially in the east. The Scandinavian ice sheet extended east to the Urals, but Siberia remained largely ice-free except in the northwest and in the mountainous northeast. A large Pleistocene refuge for temperate Asian plants was also readily available in southern China. These climatic (and related sea-level) changes, plus the physiographic complexity of

TABLE II

Global Climatic Zonation in Lowlands, with Regions in Asia

Zone	Temperature extremes	Significance	Asian examples
Tropical	No frost or other "cold" temperatures ever	Sensitivity of many tropical plants to nonfreezing cold	South Asia and East Indies
Subtropical	Occasional frost or near frost, not every year and not below about −1°C	Frost sensitivity of tropical evergreen and most other tropical plants	Southern China, Taiwan, south Asian foothills, northern Arabian Peninsula
Warm-temperate	Light to moderate frost, every year or nearly so; absolute minima not < −15°C	Annually leaf-changing broad-leaved evergreens tolerate, (sub)tropical evergreens may not	Eastern, inland SE and SW China, southern Japan; Mediterranean region
Temperate	Significant frost every year; occasional temperatures below about −15°C	Coldness tolerance limit for evergreen broad-leaved plants (e.g., internal ice formation)	Northern China and Japan, Korea, interior Asia from Mongolia to Turkey
Cool-temperate	Moderate to significant frost every year, plus cool summer; minima can be > −15°C if oceanic	Growing-season warmth marginal for many typical temperate plants, including deciduous	Hokkaido (minima < −15°C)
Boreal	Cool, short summer and long severe winter; absolute minima < −15°C, perhaps < −40°C	Growing season insufficient for most deciduous trees (exceptions: larch, birch, etc.)	Most of Siberia, northern Manchuria
Polar	Short summers below 10°C and long severe winters below 0°C; extremes < −40°C unless oceanic	Growing season too cold for wood-producing enzymes; no trees or large shrubs	North-coastal Siberia, islands of Arctic Ocean

Asia, all contributed to the development and maintenance of high biotic diversity in Asia, both in and within genera. Among vascular plants, for example, both Siberia and temperate east Asia generally show higher diversity than comparable latitudes in Europe and North America.

The most complete compilation of recent estimates of both species richness and the protection status of taxa and ecosystems is titled *Global Biodiversity* (Groombridge, 1992), from which summaries will appear throughout this article. Species richness and endemism of plants in the different countries and regions of Asia are summarized in Table III. China and Malesia are thought to have about 30,000 plant species each (most not in common), India may have 15,000 species, and tropical and subtropical Asia may have a total of at least 50,000 species. By comparison, Latin America may have as many as 85,000 species, but this includes the region from Mexico to temperate southern South America. Within Asia, China has the largest number of plant species at 30,000, followed by Indonesia (20,000), India (15,000), Malaysia (12,000) and Thailand (12,000). A special feature of East Asia is its large number of endemic plant families, many of which are monospecific, as shown in Table IV. Species richness is lowest in the dry areas of the Arabian Peninsula and mainland southwestern Asia, but rates of endemism are high in some of these drier areas. Inventories are incomplete in all areas, with only small Brunei claiming more than 20% completion. No inventory statistics at all were available for Southeast Asia, except for Thailand and Myanmar.

Estimates of species richness and endemism in four major groups of animals, but without indication of inventory status, are provided in Table V (from Groombridge, 1992). Species richness is generally lowest in the drier areas, but the number of reptiles is still high. Numbers of birds are also fairly high in mainland southwestern Asia. Other sources of biodiversity data include the United Nations Environment Programme *Global Biodiversity Assessment* (Heywood, 1995) and the various publications of the World Resources Institute.

III. REGIONALIZATION AND BIOMES OF ASIA

Global regionalization of climatic types, as related to causal mechanisms (latitude and global circulation) and

TABLE III

Plant Species Richness and Endemism in Asia[a]

Location	Angiosperms	Gymnosperms	Ferns	No. of Endemics	Percentage endemism[b]	Date[c]	Status[d]
Arabian Peninsula							
Bahrain	195	1	1	0	0	1991	1 c
Kuwait	234	1	1	0	0	1991	1 c
Oman	1,018	3	14	74	7.1	1991	1 c
Qatar	220	1	0	0	0	1991	1 c
Saudi Arabia	1,729	8	22	34	1.9	1991	2 c
United Arab Emirates	340	2	5	0	—	1991	1 c
Yemen (P.D.R.)	1,373	3	41	58	4.1	1991	2 c
Yemen (Arab Rep.)	959	1	14	77	7.9	1991	1 c
Southwest Asia (mainland)							
Afghanistan	3,500	—	—	—	[30–35]	1989–1991	2 e2
Cyprus	1,650	12	20	88	5.2	1977–1985	1 c
Iran	6,500	33	—	—	[30–35]	1989–1991	1 e2
Iraq	2,914	7	16	190	6.5	1966–1986	1 c
Israel	2,294	8	15	155	6.7	1982–1984	1 c
Jordan	2,200	6	6	—	—	1982–1985	2 c
Lebanon	2,000	12	40	—	[10]	1984–1991	2 e3
Syria	2,000	12	40	—	[10]	1984–1991	2 e3
Turkey	8,472	22	85	2,651	30.9	1988	3 e2
South Asia							
Bangladesh	5,000	—	—	—	—	1972	2 e2
Bhutan	5,446	22	—	50–100	1.4	1991	3 e1
British Indian Ocean Territory	100	1	—	0	0	1971	1 e1
India	15,000	—	1,000	5,000	31.3	1983–1984	2 e2
Maldives	260	2	15	5	1.8	1983	1 c
Nepal	6,500	23	450	315	4.5	1978–1982	2 c
Pakistan	4,917	21	—	372	7.5	1986	2 e1
Sri Lanka	2,900	—	314	900	28.0	1982–1983	2 c
Southeast Asia (mainland)							
Cambodia	—	—	—	—	—	—	—
Laos	—	—	—	—	—	—	—
Myanmar	7,000	—	—	1,071	15.3	1961	4 e2
Thailand	12,000	25	600	—	—	1979–1985	3 e2
Vietnam	—	—	—	—	—	—	—
Insular and Peninsular Southeast Asia							
Brunei	3,000	28	—	7	0.2	1990	5 e2
Indonesia	20,000	—	2,500	15,000	66.7	1991	4 e3
Malaysia	12,000	—	500	—	–	1991	3 e3
Philippines	8,000	31	900	3,500	39.3	1982–1991	3 e2
Singapore	2,000	2	166	1	0.1	1989–1991	1 e1
East Asia							
China	30,000	200	2,000	18,000	55.9	1991	3 e2
Hong Kong	1,800	4	180	25	1.3	1978–1991	2 e2
Japan	4,700	42	630	2,000	37.2	1987	1 c
Korea (North)	2,898[e]	—	—	107	14.0[e]	1976–1983	— c
Korea (South)	2,898[e]	—	—	224	14.0[e]	1976–1983	— c
Mongolia	2,272	—	—	229	10.1[a]	1984	1 c
Taiwan	2,983	20	565	—	[25]	1982–1991	1 c

[a] The data are regrouped by region from those presented in the *Global Biodiversity* report (Groombridge, 1992, p. 80) compiled by the World Conservation Monitoring Centre (WCMC). The last two columns refer to the date and status of country inventories, using codes from the original source.

[b] Percentage endemism is calculated from the data, unless in square brackets.

[c] Date is the date of the information provided to the WCMC.

[d] Status includes percentage completion (first number) and method (counted/estimated): 1, <5%; 2, 5–10%; 3, 10–15%; 4, 15–20%; 5, >20%; c, counted; e1, approximate count; e2, extrapolation, e3, estimated based on any available information and comparable floras.

[e] Data for North and South Korea are combined; no data are available for combined Yemen.

TABLE IV

Endemic Families of East Asia[a]

Polygeneric families	
Helwingiaceae	(4–5/4–6)
Carlemanniaceae	(2/4, E. Himalaya, Assam, N. Myanmar, N. Indo-China to Yunnan; 1 disjunct in Sumatra)
Podoaceae	(2/3, central Yunnan to E. Himalaya and N. Thailand)
Monogeneric families	
Hostaceae	(10–40)
Stachyuraceae	(10)
Cephalotaxaceae	(9)
Aucubaceae	(7–11, to E. Himalaya and N. Myanmar, 1 to Russian Far East)
Torricelliaceae	(2, E. Himalaya, N. Myanmar to central China)
Triplostegiaceae	(2 1 disjunct in Sulawesi and New Guinea)
Euptelaceae	(2)
Trapellaceae	(1–2, extending to Russian Far East)
Monospecific families	
Nandinaceae	
Circaeasteraceae	(not Japan, excluding *Kingdonia*)
Tetracentraceae	(not Japan)
Cercidiphyllaceae	(not Himalaya)
Euryalaceae	(not Himalaya but to India and Bangladesh)
Plagiopteraceae	(disjunct in India, Bangladesh, lower Myanmar, Thailand, Guangxi)
Pottingeriaceae	(Assam, N. Myanmar, NW Thailand to Yunnan)
Trochodendraceae	(not mainland China)
Bretschneideraceae	(China to N. Laos and N. Thailand)
Dipentodontaceae	(central China to NE India, Himalaya)
Sargentodoxaceae	(only China to N. Vietnam and N. Laos)
Rhoiptelaceae	(only SW China to N. Vietnam)
Sladentiaceae	(only SW China and adjacent Myanmar, Thailand)
Acanthochlamydaceae	(only in Hengduan mtn. to E. Tibet)
Davidiaceae	(only in China)
Ginkgoaceae	(only in China)
Eucommiaceae	(only mainland China)
Glaucidiaceae	(only in Japan)
Pteridophyllaceae	(only in Japan)
Sciadopityaceae	(only in Japan)

[a] Numbers in parentheses represent number of genera followed by number of species or number of species only (if monogeneric) (from Wu and Wu, 1996).

natural ecosystems, has been portrayed perhaps best by the widely used global system of Heinrich Walter (1985). This and the major large-area landscape types, called biomes (e.g., tropical rain forest and temperate grassland), provide a geographic framework for regionalizing the ecosystems of Asia. An attempt is made in Table VI to juxtapose the main biome types, the climate type in which they occur, the corresponding Bailey ecoregion class, and the corresponding International Union for the Conservation of Nature biogeographic provinces in Asia. Note that the "Savanna" ecoregion also includes the region of tropical moist and dry deciduous forests ("monsoon forests") that stretches across South and Southeast Asia and which corresponds to the larger tropical wet–dry climatic regions of Africa and South America, with their areas of tropical deciduous forest, woodland, and savanna. Tropical deciduous forests are not recognized in some newer classifications because they have been so completely converted into savanna. Areas of tropical deciduous forest still remain in Asia, however, especially in the Ghats Mountains of India and from the Deccan Plateau to interior Southeast Asia. Note also that one must be careful with Russian and Chinese terminologies which use terms such as "Hot/Warm Continental" and "subtropical" for temperate forest regions which have cold to severely cold winters.

TABLE V

Species Richness and Endemism of Higher Animals in Asia[a]

Region	Mammals		Birds		Reptiles		Amphibians	
	Species known	Endemic species	Species known	Endemic species	Species known	Endemic species	Species known	Endemic species
Mediterranean–Middle Asia								
Afghanistan	123	0	456	0	103	—	6	1
Cyprus	21	0	80	2	23	1	4	0
Iran	140	4	—	1	164	3	11	5
Iraq	81	1	145	1	81	—	6	0
Israel	—	2	169	0	—	—	—	0
Jordan	—	0	132	0	—	—	—	0
Lebanon	52	0	124	0	—	—	—	0
Syria	—	0	165	0	—	—	—	0
Turkey	116	0	284	0	102	5	18	2
Arabian Peninsula								
Bahrain	—	0	—	0	25	0	—	0
Kuwait	—	0	27	0	29	0	2	0
Oman	46	3	—	0	64	11	—	0
Qatar	—	0	—	0	17	0	—	0
Saudi Arabia	—	1	59	0	84	5	—	0
United Arab Emirates	0	0	—	0	37	1	—	0
Yemen	—	1	—	8	77	25	—	1
Tropical Asia								
Bangladesh	109	0	354	0	119	—	19	0
Bhutan	109	0	448	0	19	—	24	0
British Indian Ocean Territory	—	0	—	0	—	—	—	0
Brunei	155	0	359	0	44	—	76	0
Cambodia	117	0	305	0	82	—	28	0
India	317	38	969	69	389	156	206	110
Indonesia	515	165	1,519	258	511	150	270	100
Laos	173	0	481	1	66	—	37	1
Malaysia	264	14	501	4	268	—	158	39
Maldives	—	0	24	0	—	—	—	0
Myanmar	300	8	867	4	203	29	75	9
Nepal	167	1	629	1	80	—	36	7
Pakistan	151	3	476	0	143	22	17	2
Philippines	166	90	395	172	193	131	63	44
Singapore	57	1	118	0	—	—	—	0
Sri Lanka	86	12	221	20	144	75	39	19
Thailand	251	5	616	2	298	39	107	13
Vietnam	273	5	638	12	180	—	80	26
East Asia								
China	394	62	1,100	63	282	—	190	131
Hong Kong	38	0	107	0	61	0	23	2
Japan	90	29	>250	20	63	28	52	35
Korea (North)	—	0	—	0	19	1	13	0
Korea (South)	49	0	—	0	18	—	13	1
Mongolia	—	6	—	0	—	—	—	0
Taiwan	62	13	160	15	67	20	26	6

[a] As in Table III, the data are regrouped by region from those presented in the *Global Biodiversity* report (Groombridge, 1992, p. 139).

TABLE VI

Biomes, Ecoregions, and Biogeographic Provinces of Asia

Biome region[a]	Climate[b]	Ecoregions[c]	Biogeographic province[d]
Tropical Rain forest	I	Tropical rain forest	Malabar rain forest Ceylonese rain forest Bengalian rain forest Burman rain forest Indochinese rain forest South Chinese rain forest Malayan rain forest
Tropical deciduous forest (moist, dry types)	II	Savanna	Indus–Ganges monsoon forest Burman monsoon forest Thailand monsoon forest Mahanadian monsoon forest Coromandel monsoon forest Ceylonese monsoon forest Deccan thorn forest
Tropical islands		—	Laccadives Islands Maldives and Chagos Islands Cocos-Keeling and Christmas Islands Andaman and Nicobar Islands Sumatra Java Lesser Sunda Islands Celebes Borneo Philippines
Subtropical desert	III	Tropical/subtropical desert	Arabian desert Thar desert
Mediterranean scrub	IV	Mediterranean	Mediterranean sclerophyll
Laurel forest (evergreen broad-leaved)	Ve	Humid-subtropical	Chinese subtropical forest Japanese evergreen forest
Subtropical islands			Taiwan Ryukyu Islands
Summergreen forest (temperate deciduous)	VI	Hot/warm continental	Manchu-Japanese mixed forest Oriental deciduous forest West Anatolian temperate forest Kamchatkan
Temperate grassland	VII	Temperate steppe	Pontic steppe Mongolian–Manchurian steppe
Temperate desert	VIIa	Temperate desert	Anatolian–Iranian desert Iranian (subalpine) desert Turanian desert Takla Makan–Gobi desert Tibetan desert
Boreal forest	VIII	Subarctic	West Eurasian taiga East Siberian taiga
Tundra	IX	Tundra	Low Arctic tundra High Arctic tundra Arctic cold desert
Multi-zonal mountains	X	Mountains	Caucasian–Iranian highlands Altai highlands Pamir–Tien Shan highlands Hindu Kush highlands Himalayan highlands Sichuan–Yunnan highlands
Lakes systems	—	—	Aral Sea Lake Baikal

[a] Biome regions are denoted by names commonly used in regional and global treatments of terrestrial vegetation and biomes (Archibold, 1995; Walter, 1985; and the "Ecosystems of the World" book series).

[b] Climate types are those of Walter (1985): I, equatorial; II, tropical summer-rain; III, subtropical arid; IV, mediterranean; Ve, warm-temperate (humid east sides); Vm, marine west-coast (perhumid); VI, typical temperate (humid) VII, temperate continental (subhumid); VIIa, temperate arid; VIII, boreal; IX, polar; X, highlands (unclassified).

[c] Ecoregions are from Bailey (1996). See also the world map from 1989.

[d] Biogeographic provinces are from the classification by Udvardy (1975).

Using the synthesis presented in Table VI, Asia can be divided into four main regions of bioclimates and associated ecosystems:

1. Tropical Asia, including the tropical rain forest and monsoon forest regions, the tropical islands, and the various elevational belts of tropical mountains
2. the dry region of subtropical southwestern and interior Middle and Central Asia, including the subtropical and interior deserts, the temperate grasslands, the Mediterranean-type ecosystems in the southwest, and the drier mountain ranges
3. East Asia, the monsoonal but extra-tropical region of subtropical and warm-temperate evergreen broad-leaved forests ("laurel forests") and of temperate deciduous forests, with the more humid uplands and mountain ranges in these areas
4. Boreal and polar Asia, including the boreal forests of Siberia, northeasternmost China, and northern Mongolia and the tundra landscapes across the northern Siberian coastal plain and in the mountain areas of eastern Siberia.

Treatment of the various ecosystems of Asia will follow this framework, with an emphasis on vegetation since it provides the basic physical and trophic structure of ecosystems. Wetlands, coastal ecosystems, and some artificial ecosystems will be treated separately.

Estimates of vegetation cover type, percentage cover, and carbon storage, by country and region, are summarized in Table VII (Groombridge, 1992). The most extensive forest cover is in eastern to southern Asia and in Siberia, with non-forest landscapes dominating southwestern Asia. These data probably overestimate the extent of remaining forest.

Allotted space does not permit a detailed description of regional vegetation composition or even a more complete bibliography. Some general treatments, from which one can find references to more regional descriptions, include the following:

Global: Archibold (1995) and Walter (1985)
Tropical: Champion and Seth (1968), Mani (1974), and Whitmore (1984)
Dry region: Walter and Box (1983) and Zohary (1973)
East Asia: Missouri Botanical Garden (1983), Numata (1974), and Wang (1961)
Boreal/polar: Walter (1974)
Mountains: Chen *et al.* (1986), Troll (1972), and Walter (1974)

For animals, one might start with the extraordinary syntheses of Schaller (1998).

IV. TROPICAL ASIA

Tropical Asia is the region of mainly low-lying topography and warm, at least seasonally humid climates which extends from India to the Philippines and East Indies (i.e., excluding all of dry southwestern Asia). Much of the region is coastal lowland, and the highest point is on an island, at Mount Kinabalu (4110 m) in northern Borneo. The region can be divided climatically into two main zones:

1. The perhumid equatorial zone, covering most of the East Indies and southern Philippines, the Malay Peninsula, and extending along the windward southwest-facing coastlines of mainland southern Asia to southern Kerala and Sri Lanka
2. The seasonally wet–dry zone to the north, mainly across interior southern Asia from India to Vietnam and the northern Philippines, but also including the eastern Sunda Islands (6–11°S) and some other island areas, including rain shadows such as eastern and northern Sri Lanka

The natural biome of the equatorial zone is tropical rain forest and other humid tropical evergreen forests, whereas the wet–dry zone contains natural landscapes ranging from moist and dry deciduous forests to savanna and small areas of dry evergreen forest mainly in Thailand. Other ecosystems, such as swamp forests, kerang (sclerophyll scrub), and viny limestone forests, occur on particular substrates. These main characteristic vegetation types of tropical Asia and their topographic or substrate affinities are summarized in Table VIII. Characteristic vegetation belts in mountains are summarized in Table IX (Whitmore, 1984).

A. Humid Tropical Forest

Tropical rain forest is tall, multi-layered, evergreen broad-leaved forest characterized by many epiphytes, climbing plants, and a tremendous richness of tree species with dark green, thin-coriaceous leaves, many with attenuated "drip tips" for better drainage but otherwise all looking very similar. The air is constantly humid, but the canopy is exposed each day to high solar radiation, resulting in a drier canopy microclimate and a more perhumid microclimate in the forest understory. Many tree species respond to this difference with somewhat smaller, thicker canopy leaves and larger, thinner (and often darker) understory leaves, all on the same tree. In true tropical rain forest, the only seasonality is the

TABLE VII

Estimates of Vegetation Type and Percentage Cover in Asia[a]

Region	Coastal aquatic (%)	Major wetlands (%)	Desert/ semidesert (%)	Polar/ alpine (%)	Grass/ shrub (%)	Crop and settlements (%)	Interrupted woods (%)	Major forests (%)	Carbon (kg/m²)
World	4	2	13	12	20	11	17	22	3.1
Asia	4	1	16	9	24	17	10	18	2.6
Middle East									
Cyprus						100			0.8
Iran	1		30		41	7	17	4	1.4
Iraq			30		34	33	2		1.0
Israel					40	30	30		1.7
Jordan			44		49	7			0.6
Lebanon						25	75		2.5
Syria			18		41	36	3	1	1.0
Turkey	3		9		37	18	25	7	1.9
Arabian Peninsula									
Kuwait	9		91						0.3
Oman	13		46		34	7			0.5
Qatar					100				0.9
Saudi Arabia	1		62		33	2	1	1	0.5
United Arab Emirates	3		76		21				0.2
Yemen	7		11		74	3	6		1.6
USSR (former)	3	2	5	26	10	8	21	26	3.2
South Asia									
Afghanistan			11	16	62		7	4	1.2
Bangladesh	7				5	42	25	20	4.6
Bhutan				25		25	19	31	3.7
India	3	0	2	2	12	44	23	14	2.7
Nepal				23			35	42	4.4
Pakistan	1	2	21	7	46	13	9	0	1.1
Sri Lanka	32				42	3	13	10	1.7
Southeast Asia									
Brunei		67				33			2.3
Cambodia	7	4				19	5	65	6.5
Indonesia	24	9			4	9	14	40	5.4
Laos						7	7	86	9.1
Malaysia	8	3			2	10	33	45	6.7
Myanmar	5			1	2	16	23	52	6.7
Philippines	32				2	22	13	31	4.7
Thailand	7					41	7	45	5.5
Vietnam	17	1			4	30		49	6.0
East Asia									
China	1	1	14	22	21	17	5	18	2.4
Japan	21			1	5	18	27	30	4.2
Korea (North)	12				18	14	16	39	3.8
Korea (South)	10				24	20	24	22	3.4
Mongolia			32	5	52	1	3	7	1.4
Taiwan	29					12	6	53	5.5

[a] Estimates are from the *Global Biodiversity* report (Groombridge, 1992, p. 251), based mainly on Olson *et al.* (1983).

difference between day and night. As a result, the main phenological functions of the trees tend to occur simultaneously. A given tree may be seen to be flowering, flushing new leaves, dropping individual or small bunches of old leaves, and carrying ripening or ripe fruit simultaneously on different branches.

As one proceeds away from the equator, subtle climatic cues begin to appear: seasonal periods of less

TABLE VIII

Tropical Humid Forest Types of Southeastern Asia[a]

Climate	Soil water	Zone	Elevation	Soil type	Forest type
Everwet	Dry land	Inland	Lowland (<1200 m)	Zonal	Lowland rain forest
			Mountains		
			1200–1500 m		Lower montane rain forest
			1500–3000 m		Upper montane rain forest
			3000 m to treeline		Subalpine forest
			Lowland (mostly)	Podzolized sands	Heath forest
				Limestone	Limestone forest
				Ultrabasic rock	Forest on ultrabasic rock
		Coastal		Sand	Beach vegetation
	Near surface	Saltwater			Mangrove forest
		Brackish water			Brackish-water forest
		Fresh water		Oligotrophic peat	Peat swamp forest
			Wet permanently	Eutrophic (muck and mineral)	Freshwater swamp forest
			Wet periodically		Seasonal swamp forest
Wet–dry	Moderate dry season	Inland			Semievergreen forest
	Increasing seasonal deficit				Moist deciduous forest
					Dry deciduous forest

[a] From Whitmore (1975), based on Steenis (1950).

precipitation and slight seasonal temperature variations. The result is a "tropical seasonal evergreen forest" which is still tall and very diverse but in which phenological functions become synchronized and more sensitive plant types become somewhat less abundant, such as epiphytes with only aerial roots. Further from the equator, as the tropical dry season begins to become more pronounced, the forest becomes a "tropical semievergreen forest" containing some deciduous trees and still fewer epiphytes (Walter, 1985). In most tropical trees, deciduousness appears to be a facultative characteristic. This is demonstrated by two facts:

1. Many tropical tree species are evergreen in one part of their range but lose their leaves in a drier part.
2. Many individual tropical trees have been observed to keep almost all their leaves in wetter years and lose most or all of their leaves in drier years.

Temperate tree species, on the other hand, often die within a few years if planted in the tropics, apparently due to lack of the seasonal cues necessary for vernalization.

Tropical rain forest has its largest extent in Asia from peninsular Malaysia through the East Indies and

TABLE IX

Generalized Vegetation Belts in Mountains of Tropical Southeastern Asia[a]

Belt	Elevation (m)	Floristic zone	Main taxa
Upper montane	1500–2100	Montane ericaceous	Ericaceae, Myrtaceae, conifers
Lower montane	1200–1500	Oak-laurel	Fagaceae, Lauraceae
	750–1200	Upper dipterocarp	*Shorea, Dipterocarpus*
Lowland	300–750	Hill dipterocarp	Lowland + *Shorea curtisii*
	0–300	Lowland dipterocarp	Many dipterocarps, especially *Dipterocarpus, Shorea, Dryobalanops aromatica*

[a] From Whitmore (1975; 1984) for mountains of Malaya, modified from Symington (1943). The oak-laurel zone corresponds closely to subtropical/warm-temperate laurel forests further north (Ohsawa, 1995).

southern Philippines. It can also be found, however, along the mainland coast from peninsular Thailand and Myanmar to Bangladesh, in Kerala and southwestern Sri Lanka, and northward through the mountains of Southeast Asia, Myanmar, and Assam to southernmost China (Yunnan) and the foothills of the eastern Himalaya (Whitmore, 1984). In these more northerly areas, extending beyond the Tropic of Cancer, the rain forest takes on a more subtropical character but maintains high diversity and its multi-layered structure, with abundant epiphytes and lianas.

Tropical rain forest in general is composed of many tree species from families such as the Lauraceae, Rubiaceae, Moraceae, Meliaceae, and Leguminosae, the latter two usually with compound leaves and somewhat smaller leaflets than the typical rain forest leaf. Tropical rain forest reaches its greatest species richness in Southeast Asia, but here another family, the Dipterocarpaceae, provides more of the rain forest species than does any other family elsewhere. Some of the most important tree genera in Malaysia, Indonesia, and adjacent mainland areas include *Dipterocarpus, Shorea, Hopea, Dryobalanops,* and *Vatica* (Dipterocarpaceae); *Mitragyna* and other Rubiaceae; trees with large buttress roots such as *Heritiera* (Sterculiaceae); *Aglaia* and *Dysoxylum* (Meliaceae); *Mangifera* and other anacards; *Ficus* and *Artocarpus* (Moraceae); other trees with large edible fruits such as *Durio* (Bombacaceae); *Koompassia* and other legumes; and other families such as Dilleniaceae, Flacourtiaceae, and Thymeliaceae. Most genera are mainly tropical. The higher understories contain trees of the same physiognomy but also various palms (e.g., *Licuala*), even climbing palms (rattans), and other climbers such as stranglers (especially *Ficus* spp.). The ground layer is largely composed of frutescent but entirely herbaceous monocots such as *Heliconia* and *Alpinia* (both Zingiberidae). Tall bamboos, of genera such as *Calamus* and *Phyllostachys,* may form colonial stands where the forest has been disturbed frequently and opened, especially in mountains. Secondary vegetation may also involve larger-leaved, deciduous, fast-growing arborescents from the characteristic pioneer genera *Macaranga* and *Mallotus* (analogous to *Cecropia* in the tropical Americas). Some other characteristic secondary genera include *Melastoma, Glochidion* (Euphorbiaceae), and *Trema* (Ulmaceae).

Birds, amphibians such as tree frogs, reptiles such as tree-dwelling lizards, and countless insects are the main groups of animals in tropical rain forests. Mammals are less numerous but include both larger ground-dwellers such as rodents and flying mammals (i.e., bats). Most animals are probably insectivorous, although many are also mixed feeders, including various mammals which range between the canopy and the ground. Some larger ground animals are herbivores, eating mainly plant roots and fallen fruit. Ants are an especially interesting and important group for their role in dispersing plant diaspores and in decomposition and nutrient cycling.

B. Monsoon Forest (Moist Deciduous Forest)

Monsoon forest is not always a well-defined term but is used most commonly for the tall tropical "moist deciduous forest" with fairly large, soft (malacophyllous) deciduous leaves which occurs in areas with essentially perhumid rain forest conditions during a wet season which is longer than the dry season. Since the "raingreen" leaves wilt when soil water is depleted (about a month after the end of the rainy season), there are generally no bright colors in the tropical "autumn" (with the exception of *Excoecaria agallocha* in mangroves). In the tropical "spring," however, many trees bloom brightly in the month before the rainy season, cued by increasing temperatures, and the new leaves of many species are also bright reddish before enlarging and receiving their full dose of chlorophyll. Moist deciduous forest is not as species-rich as the rain forest but may still reach canopy heights of 30 m and have more species than temperate forests.

Moist deciduous forest occurs in the Western Ghats (Kerala) and in those parts of lowland and mountainous eastern India, Myanmar, and Southeast Asia which have relatively short dry seasons (Champion and Seth, 1968). In the Western Ghats, one of the most important trees is teak (*Tectona grandis*), which grows very straight but has unusually large, soft leaves very subject to insect herbivory. Other important trees include *Shorea* spp., *Terminalia tomentosa, Lagerstroemia lanceolata, Grewia tiliaefolia,* and *Pterocarpus marsupium.* To the east one encounters sal (*Shorea robusta*), a dipterocarp and another of the most important forest trees of tropical Asia, economically and ecologically. Moist deciduous forests with *Shorea* continue eastward from the Eastern Ghats through Assam and Myanmar into parts of Southeast Asia. The composition and biogeography of the moist as well as dry deciduous forests of India are well described by Mani (1974), who provides photographs of various forest areas which are now probably gone. Other moist deciduous forests occur in scattered mountain areas of Sri Lanka and the East Indies but have been largely destroyed over most of their area.

C. Tropical Dry Forest, Scrub, and Savanna

"Dry deciduous forest" occurs where the dry season is at least as long as the wet season and involves shorter, somewhat more open forests dominated by trees with usually smaller leaves, including many Leguminosae. Species richness and leaf area are lower than in monsoon forest or rain forest, and canopy height is usually not more than about 15 m. In still drier areas, dry forest may grade into dense but shorter thorn-scrub, dominated by sprawling, thorny arborescents with sparse foliage of quite small leaves (or leaflets), many of which may be Leguminosae. In other drier areas, especially on relatively flat terrain, trees may be only widely scattered in savannas dominated by tall to short seasonal grasses. In most areas, all three of these vegetation structures may be interspersed in mosaic landscapes, with forest now usually replaced by more extensive areas of degradation stages: thorn-scrub, sclerophyll scrub, and derived savanna. This kind of raingreen mosaic landscape occurs over much of the Deccan Plateau and Ganga Valley, in the "dry belt" of interior Myanmar, and in much of interior Southeast Asia. It also occurs on the drier northern to eastern side of Sri Lanka and in drier parts of the northern Philippines and the East Indies, especially in the Lesser Sunda Islands of eastern Indonesia. An interesting variant of this raingreen response to the tropical dry season is represented by the somewhat open "dry evergreen forest" of interior Thailand.

The gradual transition between moist and dry deciduous forest occurs in particular across the Deccan Plateau, from moist deciduous forest in the Western Ghats to thorn-scrub and savanna in the rain shadow immediately eastward and to dry deciduous forest, and again to moist deciduous forest further eastward. Teak and sal still occur in the dry forest, but trees such as *Diospyros melanoxylon* (ebony), *Anogeissus latifolia* (Combretaceae), leguminous taxa such as *Acacia* (especially *A. catechu*) and *Cassia,* and low arborescents such as *Zizyphus* become more important. This is also the home of the evergreen but semi-parasitic sandalwood tree (*Santalum album*). Further east, in interior Southeast Asia, the main type of dry deciduous forest is "dry dipterocarp forest," reaching over 20 m in height and with canopy species such as *Dipterocarpus tuberculatus, D. intricatus, Shorea obtusa,* and *Sh. siamensis.* In addition, dry evergreen forest dominated by evergreen dipterocarps such as *Hopea ferrea* and *Shorea henryana* covers significant areas, especially in Thailand.

D. Humid Tropical and Subtropical Mountains

The main truly tropical or subtropical humid mountain areas in Asia are the lower slopes of the Himalaya (south side), the Western and Eastern Ghats in peninsular India, the mountains of Southeast Asia, and the smaller ranges and isolated volcanic peaks of the East Indies, the Philippines, and Sri Lanka. A general zonation scheme for the vegetation of mountains in tropical Asia was shown in Table IX. As one ascends any mountain range, two climatic changes occur which together quickly reduce the effect of any dry season which may exist in the lowlands:

1. Temperature decreases (approximately 6°C per kilometer), reducing potential water loss to evapotranspiration.
2. Humidity and precipitation increase, at least up to the quasi-permanent cloud belt, which is usually 1000–1500 m above the base of the mountains (lower in more humid areas).

Thus, a tropical dry season in the surrounding lowland may disappear completely with a climb of 1000 m or less. At some elevation, generally between 800 and 1200 m, the lowland forest is replaced by a belt of tropical montane forest reminiscent of the evergreen laurel forest of the warm-temperate zone and dominated by temperate taxa such as Ericaceae and *Ilex,* plus subtropical taxa such as Myrsinaceae, Symplocaceae, Rubiaceae, and understory Lauraceae and Theaceae. From the lowland forest to this montane forest, leaf size typically changes from normal mesophyll to a mix of smaller notophyll and microphyll (Ohsawa, 1995). Above this there may be a low, dense, distinctly microphyll cloud forest of laurophyll treelets covered by mosses and other epiphytes. Of the truly tropical Asian mountains (i.e., excluding the upper parts of the Himalaya), however, only Mt. Kinabalu is high enough for an alpine belt and it is only rock (old volcanic plug) at its top.

Mount Kinabalu (4110 m, northern Borneo) is the highest mountain between the Himalaya and New Guinea (geologically part of Australia). Due to its location near the equator, Mt. Kinabalu contains taxa from both hemispheres, including some distinctly Australian elements such as *Leptospermum* and *Platycladus.* Due to the discontinuous nature of the Southeast Asian landmass and its fluctuating biogeographic history, Mt. Kinabalu also represents a true biogeographic "island in the sky" and one of the most interesting goals of any knowledgeable geobotanist. The lower slopes are covered by lowland and montane (subtropical) rain forest, followed by laurel forest rich in *Syzygium* and other amphi-tropical taxa. Near the probably edaphic treeline (around 3500 m), on shallow soil, is a belt of low, open forest dominated by *Leptospermum,* with extremely small (leptophyll) leaves reminiscent of Australian forms and the *Kunzea* woodlands of New Zealand. Other interesting features include 30-cm-wide, brown, foul-smelling, fly-pollinated flowers of *Rafflesia* and the diversity of *Nepenthes* species, with their insectivorous pitchers hanging on vines.

The mountains of mainland Southeast Asia are less described but generally show the same vertical zonation as described elsewhere (Whitmore, 1984). Evergreen *Castanopsis, Lithocarpus,* and *Cyclobalanopsis* (evergreen *Quercus*), dominants in the laurel forests of East Asia, extend far into the tropics in Southeast Asia to form extensive montane forests. Satellite data have suggested that some relatively large areas of montane rain forest and laurel forest may remain.

In the Western Ghats of Kerala, moist deciduous forest forms a lower montane forest belt, followed upward by semi-evergreen transition and then evergreen laurophyll cloud forest. The same zonation probably occurred on other mountains of southern India and Sri Lanka but has been largely destroyed, converted especially for tea plantations. A similar zonation occurs on the lower southern slopes of the eastern Himalaya but involves more subtropical species.

V. DRY REGIONS OF WEST AND INTERIOR ASIA

The dry regions of Asia include the subtropical deserts of southwestern Asia, the adjacent region of Mediterranean-type ecosystems from Turkey to Israel, the Asian portion of the temperate grassland corridor from the Ukraine to Mongolia and northern China, and the semi-desert and desert regions of both Middle Asia (Turkestan) and Central Asia (Gobi and northwest China as well as the high, dry Tibetan Plateau). Major mountain ranges separate some of these and adjacent regions, especially the Himalaya, which blocks moist monsoonal air from the south, and the Tien Shan system, which blocks moist air from the Mediterranean and thus separates summer-dry Middle Asia from summer-rain Central Asia. Most of these dry climates are also continental, reaching coastal areas only in the Middle East. The Mediterranean region contains landscapes ranging from small areas of sclerophyll forest and montane conifer forest to sclerophyll shrublands and dwarf-scrub. The temperate grasslands range from tall-grass prairies and meadow-steppes in the north to drier, open steppes in the south, following the north–south moisture gradient which obtains across most of the continent. These steppes eventually grade southward into the temperate semi-deserts and deserts.

A. Subtropical Deserts

Deserts (*sensu strictu*) have essentially no vegetation, whereas semi-deserts have scattered plants. The subtropical arid zone of Asia contains large areas of extreme desert: the Rub al-Khali (empty quarter) of the southeastern Arabian Peninsula, some areas of the Syrian desert (in Iraq and northern Arabia, not Syria), and the Dasht-e-Kavir and smaller areas of interior Iran. The rest of the subtropical desert area, from the Middle East to westernmost India, is mostly semi-desert, with widely scattered xeromorphic dwarf-shrubs, desert grasses, and ephemeral herbs on coarse, often gravelly or even rubbly substrates.

Desert conditions may have existed since the earlier Paleozoic. Whether the desert flora arose this early, or in the Miocene (expanding during the colder, drier Pliocene and Pleistocene), the desert flora of Asia has probably been closely related to that of northern Africa for a long time. Although total species richness is low and perhaps because there are essentially no true native stem-succulents, the vegetation of the Middle Eastern deserts shows a wide variety of adaptations of form and seasonality to avoid dehydration (Walter, 1985; Zohary, 1973). Some dwarf-shrubs are evergreen but greatly reduce their leaf area at the beginning of the dry season (e.g., *Artemisia monosperma, Reaumuria palestina, Salsola villosa, Suaeda palestina,* and *Zygophyllum dumosum*). Some are wintergreen shrubs, losing their leaves in summer (e.g., *Lycium arabicum* and *Anagyris foetida*), whereas some (generally dwarf-shrubs) shed their leaves but retain evergreen stems, as in *Retama* and *Calligonum* species. Some have grayish, soft, deciduous leaves but keep smaller leaves during the summer (e.g., *Artemisia herba-alba*). Some are totally leafless, such as *Ephedra.* Perhaps most interesting are the leafless shrubs with modular stems and green, photosynthetic but disposable bark segments (e.g., *Haloxylon persicum* and *Anabasis articulata*). Many of these desert dwarf-shrubs extend well into the region of cold-winter continental deserts, in Middle and even Central Asia.

Where more water is available, there may be scattered trees such as acacias, deciduous thorn-shrubs such as *Noaea,* or desert grasses such as *Aristida.* Date palms (*Phoenix dactylifera*) occur around oases, where groundwater is reliable and near the surface. If seasonal pools form, small marshes of sedges and reeds occur. Halophytes include species such as *Halocnemum strobilaceum* on very saline sites and *Nitraria retusa* in brackish water. Deciduous but otherwise juniper-like tamarisk trees (*Tamarix* spp.) also occur where groundwater can be reached, generally at greater depth.

These dry areas are the landscapes of Judaism, Christianity, and Islam. These are also the landscapes of animals such as camels, which sweat but tolerate enormous loss of body water, and the oryx and eland, which pant to moisten nasal sinuses, thus cooling the blood supplied to their brains. There are also many rodents, living in underground burrows and foraging only at night, as well as ectothermic lizards and other reptiles. Most birds spend less time in deserts but may still be dependent on desert ecosystems during the cooler season, migrating to northern Eurasia in summer.

The two main mountain ranges in the subtropical desert zone are the Hejaz (Yemen and southwestern Saudi Arabia) and the Zagros and other mountains of southern and central Iran. The montane belt of the Hejaz contains dry juniper woodlands (*Juniperus*).

B. Mediterranean Ecosystems

Mediterranean-type climates have cool winters with some precipitation and long, dry summers with no significant precipitation perhaps for several consecutive months. As a result, the primary growing season is restricted to springtime, when temperatures are increasing and soil water is readily available. Plant activity may continue into the summer as long as root systems can still reach deeper soil water. A shorter, secondary growing season often occurs in the autumn, when westerlies again bring some precipitation but before temperatures drop too low. The most characteristic plant adaptation to this climate is woody vegetation with deep root systems and hard (sclerophyll) evergreen leaves, often with waxy cuticles and/or oily secondary compounds which retard water loss from exposed surfaces. This type of "mediterranean sclerophyll" vegetation occurs in all five of the world's mediterranean-climate regions and represents a striking example of convergent evolution. The mediterranean region of Eurasia lies mainly between southern Europe and North Africa but extends eastward along the coastal rim of western and southern Turkey, to Cyprus, and eastward from Syria, Lebanon and Israel to more continental Iraq, the southern Caucasus region, and Iran.

Three general types of mediterranean landscapes can be recognized, differentiated primarily by total water availability. Where sufficient soil water can be reached throughout the year, forests composed of sclerophyll oaks such as *Quercus ilex* covered the Mediterranean borderlands (more in the more humid west) before they were cut by the ancient Romans to build their navies. Where less but still some water is available through the summer, perhaps at greater depth, the landscape is dominated by the characteristic dense sclerophyll woodlands and scrub denoted by their French name *maquis.* Where soil water is not available through the summer, more open stands of smaller, summer-deciduous dwarf shrubs dominate landscapes generically called *garrigues* (which may be patchy and may include some sclerophylls). Other plant types, however, also occur in mediterranean ecosystems on permanently more moist sites, including more water-demanding, evergreen laurel, myrtle, oleander, and even soft-leaved, winter-deciduous shrubs.

The Mediterranean region of Eurasia has been settled and modified by humans for several millennia, with consequent massive soil erosion and other degradation. As soil was lost, both soil water-holding capacity and local precipitation were reduced, resulting in replacement of forest and woodland by various types of garrigues in typical degradation stages.

In Mediterranean mountains, where winter temperatures may become too cold for sclerophylls, dry forests of pines are common, especially in Turkey. Cypresses (*Cupressus* spp.) and true cedars (*Cedrus*) may also occur, the latter occurring only in the mediterranean mountains of Lebanon (*C. libani*) and Cyprus (*C. brevifolia*), in the Atlas Mountains of northwestern Africa (*C. atlantica*), and in the drier western part of the Himalaya (*C. deodara*). Summergreen trees may also appear in mediterranean mountains and may form deciduous forests in sub-mediterranean borderlands (mainly in Europe), in continental areas with colder winters (such as the southern Caucasus and northern Iran), and on particular biotopes such as continuously moist floodplains.

Many of the familiar mediterranean plants of the more humid western Mediterranean do not extend into the Asian region to the east. In particular, the stem-evergreen "broom" form of the western Mediterranean and Atlantic Europe (genera *Genista, Cytisus,* etc.) appears not to be important in southwestern Asia. Some species, however, extend throughout the western and eastern Mediterranean areas, such as *Quercus cerris, Ceratonia siliqua, Juniperus oxycedrus, Myrtus communis,* and *Pistacia lentiscus.* Others are replaced to the east by similar species, such as *Quercus calliprinos, Qu. ithaburensis, Pistacia palestina,* and *Pinus brutia.* On the other hand, some species widespread throughout the Mediterranean region may have come originally from the east, such as Aleppo pine (*Pinus halepensis*). The vegetation of mediterranean Asia is well described by Zohary (1973, pp. 130-165).

C. Temperate Grasslands

In temperate Asia, especially in the west but also in the east, precipitation decreases southward while temperatures increase. As a result, climatic dryness increases southward and the temperate grasslands of Asia occur primarily in a rather narrow corridor that extends east–west across the continent, from the Black Sea region to northeastern China, between the boreal forest to the north and drier semi-desert and desert areas to the south. The Russian-derived term "steppe" normally means a short, discontinuous grassland in English-language ecology but is traditionally used in Asia for all grasslands. The Asian grassland zone can be divided as follows:

1. A discontinuous strip of "forest steppe" in the north (mainly *Populus* and *Betula*), analogous to the aspen "grove belt" of North America
2. A zone of tall-grass "meadow steppe" with many colorful forb species, analogous to the tall-grass prairies of North America
3. A zone of shorter, more open steppe, less developed north of the Black Sea but widening east of the Caspian and extending southward toward the desert areas of Turkestan and the Gobi

This zonation, its soils, and steppe ecosystems have been well described in the Russian ecological literature, much of which is essentially unknown in the West. In China also, scientific study of the grasslands (and other vegetation regions) goes back well before the Communist period but is unknown in the West.

The Eurasian steppe corridor is broken, largely by the Altai Mountains, into the Russian–Siberian steppes and the grasslands of Mongolia and northern China (Lavrenko as cited in Walter, 1974, pp. 163–165). In the western section, the steppe is composed especially of *Stipa* species, such as *Stipa capillata, S. lessingiana, S. stenophylla,* and other important grasses such as *Festuca sulcata, Koeleria gracilis, Dactylis glomerata, Poa botryoides* and other *Poa* spp., various *Agropyron* spp., *Bromus inermis,* and species of *Carex.* Especially characteristic of the western steppe are the many small geophytes ("ephemeroids"; Walter, 1974) from genera such as *Tulipa, Crocus, Ornithigalum, Bulbocodium,* and *Hyancinthella* as well as *Poa bulbosa.* Tall forbs in the meadow-steppe include *Anemone patens, Adonis vernalis, Salvia nutans, S. superba,* and *Delphinium confusum.* These are accompanied by some shorter shrubs and semishrubs, including the genera *Spiraea, Caragana,* and *Artemisia.* The main tree species in the forest-steppe, a wooded tall-grass prairie, are *Quercus robur, Betula verrucosa,* and *Populus tremula.*

In the mountains of Middle Asia, montane steppes occur below 2000 m and taller meadow-like grasslands above, including a park-like forest-steppe transition. These may involve grasses such as *Alopecurus pratensis* and *D. glomerata* and tall forbs, such as *Delphinium confusum, Ligularia altaica, Aquilegia sibirica,* and *Scabiosa alpestris.*

The grasslands of Central Asia, east of the Altai Mountains, involve mainly the Mongolian steppe and the grasslands of Manchuria. In the Mongolian steppe the most important grasses are *Stipa* species such as *S. krylovii,* and *Aneurolepidium chinense,* plus *K. cristata, F. sulcata,* and *P. botryoides* from the west. These are joined by xeric shrubs such as *Caragana* species and *Artemisia frigida.* The Manchurian grassland is largely a meadow-steppe dominated by *Aneurolepidium chinense,* along with *A. pseudo-agropyron* and *Stipa baicalensis,* plus forbs such as *Tanacetum sibiricum* (Compositae) and xeric shrubs such as *Artemisia sibirica* (Walter, 1974). On the lower slopes of the southern Hinggan Mountains is a forest-steppe transition to montane forests of *Larix dahurica, Betula platyphylla,* and more temperate forests of *Quercus mongolica, Tilia mongolica,* and *Pinus tabulaeformis.*

The development of the world's temperate grasslands is closely linked with populations of large ungulates, such as the well-known bison of North America. In Eurasia some of the larger animals, most highly endangered, include Przewalski's horse (*Equus przewalskii*), the Saiga antelope (*Saiga tatarica*), the Zheiran gazelle (*Gazella subgutturosa*), and the wild camel (*Camelus ferus*). The ranges of these and many smaller grassland animals also extended into the adjoining semi-desert and desert zones. Smaller animals include many rodents, such as jereboas, gerbils, and small hamsters.

D. Temperate Semi-Deserts and Deserts

The temperate deserts and semi-deserts of Asia, with cold winters, occur south of the temperate grasslands in three general areas (Walter and Box, 1983):

1. In Middle Asia (Turkestan), from the Caspian Sea (below sea level) to the foothills of the Altai and Tien Shan mountains
2. In Central Asia, including Dzungaria (northern Xinjiang), southern Mongolia (Gobi), part of Chinese Inner Mongolia (e.g., Ordos desert), the Tarim Basin (southern Xinjiang), and the smaller Tsaidam Basin (Qinghai), all approximately 500–1500 m above sea level
3. On the high, cold Tibetan Plateau, with an average elevation of approximately 4000 m

Separating Middle and Central Asia is the Tien Shan mountain system, running generally from southwest to northeast. This complex mountain system, with the associated Pamir to the south and Altai to the north, rises from the dry continental interior and constitutes to some extent a multi-zonal orobiome of its own (Walter and Box, 1983), with well developed coniferous forests related more to those of humid eastern Asia.

The climates of the temperate deserts are warm to hot in the summer (except for Tibet) and severely cold in the winter, except in the southern parts of Middle Asia (Turkmenistan), near the Iranian border. Precipitation in Middle Asia occurs mainly in the spring and autumn, with very scant snow cover in winter and essentially no precipitation in summer. In Central Asia precipitation is confined almost completely to summer, brought by monsoon storms which manage to penetrate into the continental interior. Soils in all regions are coarse and skeletal, salinized in areas of internal drainage. The Central Asian deserts are totally continental and separated from the subtropical deserts of southwestern Asia. Although the Middle Asian deserts share some species with the subtropical deserts, they also are distinct from the subtropical deserts, both climatically and physiographically, separated by the Kopet Dagh and other mountains of Iran and Afghanistan.

1. Middle Asian Deserts

A characteristic of the sandy deserts of Middle Asia is the dominance of both small, typical and much larger xeromorphic shrubs, some leafless and some at least partly evergreen. Reaching 2 m in height, leafless *Haloxylon persicum* extends over the whole region and even into the Gobi desert. It is joined in the Kara-Kum desert by *H. ammodendron,* a leafless tree form reaching 8 m in some locations. This is the only cold-winter desert region in the world in which a tree form gains even local dominance. Other important desert shrubs include *Artemisia pauciflora, A. terrae-albae, Halostachys caspica, Anabasis salsa, Halocnemum strobilaceum,* and *Kochia prostrata,* along with steppe and desert grasses such as *Festuca sulcata, Stipa capillata, Koeleria cristata,* and *Agropyron repens.* In areas of flat terrain and finer soil, mini-geophytes such as *Poa bulbosa* form vernal carpets of miniature flowers (Walter, 1974, p. 254).

2. Dzungaria, Gobi, and Tarim Basin

Dzungaria is the region east of the Dzungarian Gate, a gap in the Tien Shan–Altai mountain system, and extends across northern Xinjiang. This corridor between Middle Asia and the Gobi, as well as the Gobi itself, represents a region of rockier substrates and deserts dominated by typical desert shrubs such as *Artemisia, Anabasis,* and *Calligonum.* The most important grasses are *Stipa* bunch-grasses. Isolated trees of *Populus diversifolia, Tamarix ramosissima,* and occasionally *H. ammodendron* may occur along streams.

The Takla Makan (Tarim Basin), on the other hand, is a large sand sea within the horseshoe (open to the east) formed by the Tien Shan on the north, the Pamirs in the center, and the Karakoram and Kunlun ranges on the south. Together they form one of the most effective rain shadows in the world, and precipitation has very rarely ever been recorded in the Tarim Basin. There is little vegetation on the shifting sand areas, but runoff from the mountains does generate intermittent and permanent streams which extend into the basin. Along these streams one can find riparian strips of the phreatophytic *Tamarix ramosissima* as well as the widespread *Populus diversifolia* and *Ulmus pumila.* In some areas there are even extensive marshes of *Phragmites, Typha,* and *Scirpus* species, which are especially important for the large numbers of birds which overwinter in India and migrate to Siberia for the summer.

3. Tibetan Plateau

The high, dry Tibetan Plateau (average elevation approximately 4000 m) is a region of dry steppes, alpine mats (especially of *Kobresia* species), dwarf conifer scrub of *Juniperus* and *Sabinia,* and other sparse, low-growing, often cushion-shrub vegetation adapted to the harsh conditions. Many familiar genera from Central Asia (and Middle Asia) are represented, including *Caragana, Reaumuria, Acantholimon, Astragalus, Tanacetum, Artemisia, Ptilagrostis,* and *Festuca.* Some familiar species also extend into the Tibetan highlands, including *Eurotia ceratoides* and *Kochia prostrata.* East Asian elements are important only in the east but include *Gentiana, Primula, Saxifraga, Saussurea,* and *Rhododendron.* Holarctic elements are few, but some are very abundant, especially in the alpine mats and marshes: *Kobresia, Carex, Eleocharis, Eriophorum,* and *Juncus.* Where moisture permits, subalpine tall-forb stands include such genera as *Aconitum, Delphinium, Ligularia, Polygonum,* and *Rheum.* These different geoelements all represent taxa which migrated into Tibet after the last glaciations.

VI. HUMID MONSOON ASIA (EXTRATROPICAL)

Monsoon Asia is the region dominated by the seasonally alternating monsoon wind system, which brings wet oceanic air masses onshore in the summer and blows dry and cold outward from the continental interior in winter (mainly eastward and southward). The region lies east of the Altai–Tien Shan mountains and extends from southern Siberia, through China, to Korea and Japan in the east and to tropical Asia in the south. In most of this region the only major mountains run east–west, so mean winter temperatures are low but outbreaks of even colder Siberian air are largely excluded. Except toward the interior, summer brings adequate rainfall throughout for forest growth. In fact, in Japan and some parts of eastern China, there is rarely any climatic drought and forests can be very mesomorphic. Humid monsoon Asia (i.e., forested East Asia) can be divided into two regions: (i) temperate deciduous forest in the typical-temperate and cool-temperate areas to the north and (ii) evergreen broad-leaved "laurel" forest in the warm-temperate and humid-subtropical zones to the south. Pine forests are important secondary landscapes in both areas. Mountains in East Asia are especially interesting for the rich variety of endemic conifer taxa which they contain.

A. Laurel Forest (Evergreen Broad-Leaved Forest)

Laurel forests are evergreen broad-leaved forests of warm-temperate and humid-subtropical climates, dominated by trees with intermediate-sized, dark green (shade-tolerant), thin-coriaceous but mesomorphic leaves (laurophylls) such as are especially characteristic of the laurel family (Lauraceae). Forests dominated by laurophyll trees are rather dark and somber, with low light levels below the canopy, and are evergreen from top to bottom. In some respects laurel forests can be thought of as an extension of the tropical rain forests into the warm-temperate zone. East Asia contains the world's largest area of such forests, due to its abundant rainfall and winters without severely low temperatures, at least in coastal areas or on islands. Laurel forests occur generally south of about 35°N latitude but extend to 38°N on both sides of Japan. Counterparts occur in southern Brazil and small areas of eastern Australia, as well as moist depressions ("bay forests") in the southeastern United States. Cool-temperate analogs occur in New Zealand and Tasmania.

In East Asia, laurel forests occur in southeastern China, in drier southwestern China (Sichuan and Yunnan, with different but largely vicariant species), in southern Japan, and in a small strip across southernmost Korea (Wang, 1961). In Southeast Asia, laurel forests also ascend into the mountains to form montane forests just above the tropical seasonal evergreen forests of the lowlands (Whitmore, 1984; Ohsawa, 1995). The main canopy tree genera of laurel forest are *Persea* (= *Machilus*), *Cinnamomum, Beilschmiedia,* etc. (Lauraceae); *Castanopsis,* evergreen *Quercus* (= *Cyclobalanopsis*), and *Lithocarpus* (Fagaceae); *Schima* (Theaceae), *Ilex* (Aquifoliaceae), *Michelia* (Magnoliaceae), and others, all with very similar laurophyll physiognomy. Understory trees and arborescents are largely from the same families. Lauraceae tend to be especially important in more perhumid (e.g., coastal) areas, whereas evergreen Fagaceae tend to become more important inland, especially in southwestern China.

In humid East Asia, the largest species turnover between the polar region and the tropics occurs within the bioclimatic zone of evergreen broad-leaved forests—without major change in the forest physiognomy. This occurs in the Okinawa Islands of Japan and in southeastern China, as temperate species abruptly disappear, including many laurophyll tree species, and are replaced by an essentially tropical flora which includes many new tree genera with essentially the same evergreen broad-leaved structure.

B. Temperate Deciduous Forest

The temperate deciduous forests of East Asia cover the northern half of Japan (including Hokkaido), most of Korea, and the northeastern part of China, especially Manchuria. This region is smaller than the region of evergreen broad-leaved forest and smaller than the deciduous forest region in eastern North America. Field estimates of primary productivity in Japan suggested that the deciduous forests were less productive than in North America and less productive than some montane conifer forests of Japan. Nevertheless, the deciduous forests of East Asia appear to have more species and genera, including almost all genera found in eastern

North America (except *Carya*) and in Europe, plus some additional ones (e.g., *Cercidiphyllum*).

1. Japan and Korea

The deciduous forests of Japan are especially rich, due perhaps to the perhumid maritime climate and total lack of water stress during the growing season. Deciduous forests cover most of northern Honshu, dominated mainly by *Fagus crenata* but with many other taxa, such as *Fraxinus lanuginosa, Tilia japonica,* and *Magnolia obovata.* Understories are rich with maples (e.g., *Acer mono, A. japonicum,* and *A. palmatum*) and other mesic understory elements such as *Styrax obassia, Sorbus alnifolia, Cercidiphyllum japonicum, Aesculus turbinata, Callicarpa japonica, Hamamelis obtusata, Viburnum,* and *Euonymus.* One interesting transitional forest type occurs in parts of northeastern Honshu and involves codominance by *F. crenata* and *Abies firma,* a temperate-zone fir. Forests of *Fagus japonica* generally occur only in the areas of transition to warm-temperate forest. An especially characteristic feature of the deciduous *Fagus* forests of Japan is the prevalence of short but broad-leaved bamboos of the genus *Sasa* (and recently also *Sasamorpha*), which form dense understories 0.5–1 m high where there is continuous snow cover. Other common herb-layer taxa include *Disporum, Viola, Chimaphila, Mitchella, Carex,* and ferns. Floodplain and other especially moist forests are dominated by ashes (*Fraxinus sieboldiana* and *F. spaethiana*), elms (e.g., *Ulmus davidiana*), *Pterocarya rhoifolia,* and other typical floodplain taxa (e.g., *Alnus hirsuta*). Also belonging to the temperate zone are the hemlock (*Tsuga sieboldii*) forests which occur most commonly on rocky substrates at the base of lower mountain slopes. Drier areas (still without water deficit) are generally covered by deciduous oak forests, including interior areas, lower mountains (e.g., *Qu. serrata*), and most of Hokkaido (*Quercus mongolica* var. *grosseserrata*).

The deciduous forests of Korea occur in a distinctly more continental climate, in which oaks (e.g., *Quercus mongolica*), and also *Carpinus,* become more important. Much of the Korean landscape is still recovering from the Korean War, after which the southern part was seeded with North American pines. In deciduous forest areas of both Japan and Korea, as in the evergreen forest region, native pines such as *Pinus densiflora* play a very important role in successional landscapes.

2. Manchuria and Eastern China

The forests of northeastern China and the Amur valley have a shorter growing season than in Japan but are still very rich. Canopy trees may include such species as *Quercus mongolica, Fraxinus mandschurica, Tilia amurensis* and *T. mandschurica, Ulmus macrocarpa* and *U. davidiana, Phellodendron amurense, Maackia amurensis, Juglans mandschurica, Acer mono, Betula dahurica* and *B. costata,* and *Populus koreana* and *P. ussuriensis.* A common temperate conifer in these forests is *Pinus koraiensis.* Understories may include trees of mainly the same species plus shrubs such as *Corylus heterophylla, Deutzia amurensis, Euonymus pauciflora, Philadelphus schrenkii, Syringa amurensis,* and *Viburnum* spp.

Further south, in northern, eastern, and into southeastern China, the deciduous and evergreen broad-leaved forests are almost totally destroyed by millennia of human habitation. As a result, it is difficult even for local scientists to identify the boundaries of the two forest regions. In hill and mountain areas of northern China (Liaoning, Shandong, Hebei, and Beijing) remaining mature forest types are dominated by *Quercus mongolica, Qu. liaotungensis,* and other *Quercus* species along with *Celtis, Fraxinus, Tilia, Betula,* and *Ulmus* as subordinate canopy elements. Understories may have elements such as *Lindera obtusiloba, Corylus heterophylla, Lespedeza bicolor,* and *Spiraea trilobata. Betula* and *Populus* forests occur as secondary types. *Quercus variabilis* forests occur as transitional forests south of the main deciduous forest region.

Among the most interesting deciduous forests are those involving various species of beech (*Fagus*), generally covering small areas at elevations from 1300 to 2200 m in central eastern China (Zhejiang to Guizhou provinces). Deciduous trees in interior southern China may have only short leafless periods. This is also one of very few areas in the world in which deciduous and evergreen broad-leaved trees coexist in semi-evergreen broad-leaved forests.

C. Pine Forests

Pine forests represent extensive landscapes in East Asia due to long histories of continuing forest cutting. Until after World War II, many of the evergreen broad-leaved forest areas were managed as coppice forests. Other areas, however, were burned and cleared much earlier. In hilly and mountainous terrain, soil erosion followed clearing and broad-leaved forests could no longer recover. These areas usually became pine forest and were maintained by intermittent natural and induced burning of the understory. Much of southern Japan, the Korean peninsula, and upland warm-temperate and subtropical China are now covered by such pine forests.

In Japan, Korea, and some parts of northeastern China, these secondary pine forests are mainly dominated by *Pinus densiflora*. In eastern and southeastern China, pines with more warm-temperate to subtropical distributions become dominant, especially *P. massoniana* and *P. tabulaeformis*. In southwestern China, forests of *P. armandii* occur in Sichuan and on the Yunnan plateau, whereas subtropical pine forests of *P. kesiya* (cf. *P. khasya*) become important in the mountainous transition southward into lowland Xishuangbanna and the mountains of northern Thailand. In addition to pine, the warm-temperate to subtropical "Chinese fir" (*Cunninghamia lanceolata*) is also important in secondary landscapes and, like pine, is widely planted.

D. Mountains of Humid East Asia

The mountains of East Asia represent the only large mountainous area in the warm-temperate and humid subtropical zones of either hemisphere (although smaller areas do occur in Mexico, southern Brazil, eastern Australia, and northern New Zealand). The various mountain systems of humid East Asia, including at least the more humid eastern Himalaya, thus represent areas of unusual diversity and endemism, especially among what appear to be relict conifers. Endemic conifer genera include *Amentotaxus*, *Cathaya*, *Cephalotaxus*, *Keteleeria*, *Cryptomeria*, *Cunninghamia*, *Fokienia*, *Glyptostrobus*, *Sciadopitys*, *Taiwania*, and *Thujopsis*, at least some of which once occurred more widely. In addition, more widespread conifer genera such as *Abies*, *Chamaecyparis*, *Picea*, *Taxus*, *Thuja*, and *Tsuga* have relatively high species diversity in these areas. In most of the area, the lowland biome is evergreen broad-leaved forest, which may extend well into the montane belt in most places, with well-developed *Rhododendron* belts and a more warm-temperate character. As winter becomes colder upward, summergreen deciduous forest may occur, but the forest may also grade directly into evergreen high-mountain conifer forest. None of these warmer mountain systems is high enough to have an alpine belt except the Himalaya.

In the western Himalaya, montane conifer forests are dominated by *Pinus roxburghiana* and the deodar cedar (*Cedrus deodara*), a preferred ornamental tree throughout the world. To the northwest, in the Tien Shan and Altai systems, the montane conifer forests are simpler, dominated often by *Picea schrenkiana* but including forests of shade-tolerant *Pinus sibirica* as well as boreal *Picea* and *Abies*, larches, and *Pinus sylvestris* (Walter, 1974; Walter and Box, 1983, Chap. 7). In western Tibet, the main conifers are *Abies delavayi* and *A. webbia*, along with *Picea likiangensis*. This is also the area of the dawn redwood (*Metasequoia glyptostroboides*), first identified from fossils, only discovered in the wild in the early 1900s but now widely planted as a street tree throughout China.

Extensive evergreen oak forests, involving *Quercus semicarpifolia* and many other species, still occur on the southern slopes of the Himalaya and may represent a somewhat unique warm-temperate forest biome. At higher elevations, *Betula albosinensis*, *Acer tetramerum*, and many other deciduous species form belts of summergreen forest in some areas. Subalpine conifer forests may contain species such as *Picea likiangensis* (better known from Manchuria), *Abies squamata* and *A. georgei*, *Tsuga dumosa*, and even the deciduous *Larix potaninii*. The best developed *Rhododendron* belts are in the eastern Himalaya.

On the lower parts of the eastern Tibetan Plateau, *P. likiangensis*, *A. squamata*, and *A. georgei* form tall coniferous forests, followed upward by *Abies faxoniana*, *Picea asperata*, and *P. noveitchii* (Archibold, 1995). *Larix* may be admixed, and *Sabina* and *Cupressus* form shorter woodlands in drier areas. Toward the Yunnan Plateau to the southeast, *Tsuga yunnanensis* becomes important, with *Picea brachytyla* and *Abies fabri* at higher elevation. At lower elevation, the eastern Tibetan Plateau (mainly in Sichuan) also contains two of China's biological treasures: the range of the Giant Panda at the Wolong Preserve and E-mei Shan (3099 m), one of China's four sacred mountains, with a well-developed zonation (and footpath) from laurel forest to subalpine conifer forest of *A. fabri*. To the south, at elevations not higher than about 1500 m on the Yunnan Plateau, smaller areas of laurel forest remain in the Ailao Shan and other ranges southward into Xishuangbanna.

In the transition region of eastern China (summergreen to evergreen laurel forest), two other accessible mountains also have well-developed zonations, which are botanically well described: Huang-Shan in southeastern Anhui province and Tianmu-Shan in northern Zhejiang province. The mountains of Taiwan are especially steep, and the montane laurel forests are consequently relatively well preserved.

In northeastern China, along the Korean border, the vertical zonation of the Changbai-Shan can be described as follows:

0–500 m: summergreen deciduous forest
500–1100 m: mixed deciduous and conifer forest (*P. koraiensis* plus much secondary forest of *Betula platyphylla*)
1100–1700 m: coniferous forest (*Picea jezoensis* var. *komarovii*, *P. koraiensis*, with *Abies nephrolepis*, *Acer*, *Betula*, *Sorbus*)

1700–2000 m: Birch woodland and krummholz (*Betula ermanii*)
2000–2600 m: Alpine tundra, mostly heath of *Vaccinium uliginosum* and *Dryas octopetala*, plus *Rhododendron aureum*, *Rh. confertissimum*, etc.
2600–2700 m: Alpine cold-desert (snow still just melting in early September)

Further to the north, larch forests dominated by *Larix gmelini* occur in the northern Da Hinggan Mountains near the Siberian border (Missouri Botanical Garden, 1983; Wang, 1961).

Montane forests and zonation in Korea can perhaps best be seen on Mt. Seolag (1308 m). The zonation is similar to that described for Changbai-Shan but with a greater importance of deciduous *Quercus* and *Carpinus* forests at low elevation and the lack of an alpine belt.

The mountain vegetation of Japan has been very extensively studied and described in the evergreen as well as the deciduous forest zones (Numata, 1974). Among the most impressive native coniferous forests of Japan are the tall forests of *Cryptomeria japonica* occurring on Yakushima island and in some areas with more than 3000 mm of precipitation on Shikoku and Honshu. These forests could be called temperate rain forests. Forests of *Chamaecyparis obtusifolia* are also tall and impressive. Other montane conifers in Japan include *Abies homolepis*, *A. mariesii*, and *A. veitchii* mainly on Honshu; *A. sachalinensis* mainly on Hokkaido; *Picea jezoensis* (Hokkaido) and *P. glehnii*; *Larix leptolepis* (on old lava); *Tsuga sieboldii* (lower mountains) and subalpine *Ts. diversifolia*; *Thuja standishii*; and *Thujopsis dolobrata* (endemic genus).

In most of northeastern Asia, the subalpine belt is characterized by dense stands of the krummholz species *Pinus pumila,* which also occurs in Siberia. In more maritime areas approaching the boreal zone, such as Hokkaido, conifers may be replaced by forests of birch, especially *B. ermanii.*

VII. BOREAL AND POLAR ASIA

Boreal and polar Asia extends from the Ural Mountains to the Pacific Ocean and from islands in the Arctic Ocean to the transition to temperate climates in the south, approximately along the southern boundary of Siberia. In Europe, this transition takes the form of mixed boreo-nemoral forests of deciduous broad-leaved trees and conifers, narrowing eastward from Poland through Moscow and on to the Urals around Sverdlovsk. In Asia, this transition is a more discontinuous zone of mixed forests and forest-steppe, running from the southern Urals past the southern end of Lake Baikal and on to northeastern China, where it widens again between the Siberian larch taiga to the north and the deciduous forests of Manchuria. In the middle section, the transition is mainly a narrow belt of aspen (*Populus tremula*) and birch parklands between subboreal or boreal forests to the north and the grasslands of Mongolia to the south.

During the Pleistocene, much of Siberia remained unglaciated but covered by tundra and cold steppe vegetation, which extended southward to the grasslands with no intervening coniferous forest zone. The boreal forest which now covers most of Siberia is thus a relatively recent phenomenon. This boreal forest is mainly an evergreen coniferous forest in the west but opens to lighter larch (*Larix*) forest and woodland in the east, where winters are most severe. The west-Siberian lowland is a vast complex of boreal forest and wetlands, whereas northeastern Siberia is quite mountainous. The larch region of interior eastern Siberia contains the most continental climates in the world, with mean July temperatures usually exceeding 20°C but with mean January temperatures generally near or below −30°C, reaching −51°C at Oimyakon and Verkhoyansk in the northeast. The transition to polar tundra involves a zone of open forest–tundra woodland, which reaches its farthest point north (in the world) in the Taimyr Peninsula of northwestern Siberia. Polar tundra forms a narrow belt along the northern coastal plain completely across Eurasia.

A. Boreal Forest

If the subboreal zone is included, four main types of boreal forest can be recognized (Walter, 1974):

1. Mixed forests of summergreen broad-leaved trees and conifers (mainly *Picea* and *Pinus*), along with localized pine forests
2. Dense forests of spruce (*Picea*), fir (*Abies*), and five-needle pine (*Pinus sibirica*)
3. More open forests and woodlands of Scots pine (*Pinus sylvestris*)
4. Generally open forests and woodlands of larch (*Larix*), a genus of deciduous conifers

Of these, the dense forests of spruce, fir, and Siberian pine are generally called “dark taiga,” whereas the more open forests and woodlands of pine and larch are called “light taiga.”

Boreo-nemoral mixed forests, with *Acer*, *Tilia*, *Ulmus*, *Fraxinus*, *Quercus*, *Picea*, *Pinus*, etc., are better developed in eastern Europe than in Asia but reappear in Manchuria and the Russian Far East and some smaller areas such as the subboreal region on the west side of Lake Baikal. In more continental areas, *Betula* and *Populus* may be the only important broad-leaved trees, and pines may be the main conifers, especially if the soil is coarse and less humic. At the southern end of Lake Baikal (52–54°N, 450 m), boreo-nemoral mixed forests can be found in moist alluvial lowlands, whereas dark taiga, with larch and five-needle pine, occurs on uplands higher than about 700 m.

Dark taiga is composed of shade-tolerant conifers, mainly Siberian spruce (*Picea obovata*), five-needle Siberian pine (*P. sibirica*), and Siberian fir (*Abies sibirica*). All three of these extend from the Urals into southeastern Siberia (if not completely to the Pacific). Dark taiga is limited geographically by the extreme winter cold, yielding to deciduous larch taiga generally where mean January temperature falls below about −30°C (extremes to below −60°C). The range of dark taiga is thus mainly European Russia through western Siberia and in a narrowing zone into southeastern Siberia. Within this range, dark taiga occurs over most of the area, with more open forests of *P. sylvestris* confined mainly to sandy, upland rock, or other drier substrates. In some smaller areas, mainly in mountains, other species of *Picea* and *Abies* appear as well as other five-needle pines such as *P. cembra* and *P. pumila*, often occurring as subalpine krummholz. Larches, often large, also occur widely as companion species within the dark taiga.

Scots pine (*P. sylvestris*) has the widest distribution of any tree species in Eurasia and covers large areas of light taiga throughout the European and Asian boreal zone and occurs in mixed forests in Europe, the Caucasus, and northern Manchuria. Boreal pine forests are usually open and have understories of two main types: *Cladonia* and similar lichens (plus mosses) on drier sites such as sand, and ericads such as *Vaccinium* on more moist sites.

Most of the boreal zone of central and eastern Siberia is covered by extensive open forests and woodlands of larch, especially *Larix dahurica* but also other species with smaller ranges. *Larix sibirica* is more important as a companion in the dark taiga of western Siberia. As in many higher latitude mountainous regions, the larch woods may occur mainly on slopes while flat valley bottoms remain largely treeless, a pattern attributed to cold-air drainage. The ground vegetation in these areas and in the larch stands as well is composed largely of boreal and tundra dwarf-shrubs such as *Vaccinium uliginosum*, *Ledum*, *Dryas*, and *Arctous*, plus abundant mosses and lichens. In moist areas shrubs from the genera *Juniperus*, *Salix*, *Alnus*, and *Betula* often become important understory elements.

The larch woods of interior northeastern Siberia in particular occur where annual precipitation is generally under 300 mm and sometimes under 200 mm. Although the growing season is thus a period of climatic water deficit, the trees survive from soil water which is effectively rationed to them over the course of the summer by the slow thawing of the soil (Walter, 1974, p. 84). Larches are outcompeted by evergreen conifers in some boreal areas but have no tree competitors in the coldest parts of eastern Siberia. Larch-dominated landscapes thus extend from the polar treeline across much of Siberia southward to Mongolia and into inland areas of the Russian Far East (Primorye and Khabarovsk provinces) and mountains of northernmost Manchuria.

Mountains within the boreal zone may quickly reach treeline. On lower slopes, however, the boreal forest may be somewhat different from the surrounding flatter lowlands. In the Ural Mountains, forests on the wetter western slopes generally involve *P. sylvestris*, *Abies sibirica*, *Picea abies*, *Betula pubescens*, and krummholz of *P. pumila*, whereas forests on the drier eastern slopes are mainly *P. sylvestris*. In less extreme, maritime areas of eastern Siberia, the deciduous larch taiga may be replaced again by evergreen conifers, such as *Abies nephrolepis* around the Sea of Okhotsk and *A. sachalinensis* on Sakhalin island (also northern Japan). Stunted forests of *B. ermanii* appear in the most maritime areas (e.g., Kamchatka, Sakhalin, and northern Japan).

B. Polar and Mountain Tundra

Polar ecosystems, generally called tundra, occur across the northern coastal plain of Eurasia but also in both upland and high-mountain situations, such as the lower mountains of boreal northeastern Siberia and the high Tibetan Plateau. Upland tundra in the boreal zone represents the first upward vegetation belt and is thus commonly called montane tundra. In high mountains, tundra-like ecosystems occur above the alpine treeline and are generally called alpine tundra. In all three situations, these are treeless landscapes dominated by mosaics of the following basic associations:

1. Largely evergreen, microphyllous dwarf-scrub on locally somewhat higher, drier areas
2. Wet graminoid vegetation in the lowest areas, generally flat, marshy depressions

3. Mixes of evergreen and summergreen plants, herbaceous and dwarf-shrub, including minigeophytes, on the broad, slight slopes between the higher and lower areas
4. Mainly mosses and lichens on the most extreme areas, including the coldest but also the most exposed upland areas where little snow remains

In mountains, two additional types can be identified. One is the "snow valleys" (Schneetälchen), depressions with deep snow accumulations which have shorter snow-free seasons and specialized floras composed of plants which can complete their life cycles in periods as short as one month in summer. The other is the extensive alpine mats, often dominated by *Kobresia*, found in drier high-mountain areas such as the Tien Shan, Hindu Kush, and Tibetan highlands, generally above 3000 m.

The polar tundra extends from extreme northern Fennoscandia across northern Russia to coastal northeastern Siberia in a narrow strip which is widest on the Taimyr peninsula. Due to continentality effects, however, the Taimyr is also the area where boreal forest extends farthest north, reaching about 72.5°N (as larch forest) north of Khatanga. On the flat coastal topography, herbaceous tundra is perhaps most extensive. The polar tundra can be represented zonally, however, in four subzones from south to north:

1. The transitional forest–tundra of largely dwarf-shrub tundra with scattered but still tall individual conifers (mainly larches in the east and spruces in the west)
2. Dwarf-shrub tundra, dominated largely by shrub forms from summergreen tree genera, mainly *Betula*, *Salix*, and *Alnus*
3. Largely herbaceous tundra with few dwarf-shrubs but still a continuous vegetation cover including many mosses and lichens
4. The discontinuous High Arctic cold-desert with vegetation restricted to more scattered individuals of a few vascular plants (especially *Dryas* spp.) and various lichens and mosses

The forest–tundra and dwarf-shrub tundra zones of the lowlands are better represented in western Siberia, but montane analogs are fairly widespread in the lower mountain areas east of the Lena River in eastern Siberia. Cold-desert, on the other hand, is a bit more widespread in the east and on the Arctic islands.

Two other prominent features of the polar zone are the prevalence of permafrost and substrates divided into stone polygons by frost action. Permafrost occurs where mean annual temperatures are below 0°C, i.e., north of about 62°N continuously and discontinuously in uplands further south. Polygon substrates (and other features such as pingos) occur where seasonal freezing and thawing of "soil" which is largely water forces the larger solids (i.e., rocks) to the surface in a slow, bubble-like fashion that distributes them into polygons which become connected into polygon networks. The interiors of the polygons are lower, flat, and wetter, whereas the rocky borders, only 10–20 cm higher, represent drier biotopes.

Polar tundra in particular (including boreal montane tundra) involves many taxa with circumpolar distributions, including genera such as *Carex*, cotton grass (*Eriophorum*), and other sedges; many ericads (e.g., *Vaccinium*, *Arctostaphylos*, and *Cassiope*); other forbs, such as *Dryas*; and both evergreen dwarf-shrubs (e.g., *Empetrum*) and sometimes taller deciduous dwarf-shrubs (e.g., *Betula*, *Salix*, and *Alnus*). For nonvascular plants, an even larger percentage of the taxa are circumpolar, including mosses such as *Sphagnum*, *Polytrichum*, and *Hylocomnium*, and lichens such as *Cladonia* and *Cladina*. As in the boreal zone, the flora of eastern polar Siberia is richer than that of western Siberia due to its lack of Pleistocene glaciation.

Most of the same genera, of vascular and nonvascular plants, also occur in the high-mountain tundra. Alpine treeline in mountains of northern Asia is commonly formed by *P. pumila* krummholz, whereas the alpine tundra typically involves familiar heath taxa such as *Empetrum*, *Phyllodoce*, *Cassiope*, and *Dryas* as well as dwarf *Salix* and sedges. In general, however, alpine vegetation is more species rich than is corresponding polar tundra, perhaps due to the lack of permafrost and the greater number of microhabitats in the more heterogeneous terrain of high mountains. Many more localized species usually occur. Some important genera of Asian high mountains which are not important (if present at all) in polar tundra include *Geum*, *Gentiana*, and *Leontopodium*.

Despite its relatively low productivity, the polar tundra supports many animal species and in surprisingly large numbers, at least in summer. Caribou (reindeer) winter in the boreal zone and migrate to the tundra during the summer. Birds migrate from much greater distances away, some from as far as tropical Asia. Other birds, such as ptarmigan, overwinter near the polar zone. Mammals such as polar bears and some small rodents (e.g., lemmings) also live year-round in the polar zone.

VIII. WETLANDS AND COASTAL ECOSYSTEMS

Wetlands include mangroves, other coastal ecosystems such as strand forests, beach vegetation, salt marshes, deltas, and estuaries and terrestrial wetlands such as swamp forests, marshes, bogs, and (*sensu lato*) even lakes and streams. Most wetlands are highly productive systems when not badly degraded, have important economic and social benefits for local human populations, and perform useful ecological services such as cleansing wastewater. Wetlands also support a wide range of animal life, often far beyond their own borders, as in the case of migratory birds, fish, and shellfish.

Mangroves are salt-tolerant forests growing in the intertidal zone, thus flushed by saltwater tides once or twice daily but also influenced by the inflow regime of freshwater runoff from the land. Behind many mangroves are so-called "back mangroves" which are less influenced by wave energy, generally less salty, and contain different and usually more species. Mangroves have developed in part by trapping fine sediment, so the mangrove substrate is usually deep mud (but also shallower mud and sand). As a result, mangroves are home to abundant burrowing shrimp, crabs, and other crustaceans as well as spiders, insects, and some birds. The total number of plant "mangrove species" in the world has generally been estimated at fewer than 100, involving especially the genera *Rhizophora* and *Avicennia* as well as particularly Asian genera such as *Bruguiera*, *Ceriops* and *Kandelia* (Rhizophoraceae), *Sonneratia*, and *Aegiceras*. Back mangroves contain additional taxa such as *Lumnitzera*, *Xylocarpus*, the mangrove palm (*Nypa fruticans*), *Acanthus*, the widespread fern *Acrostichum aureum*, and even the buttressed tree *Heritiera littoralis* (Sterculiaceae) in some cases. In areas protected from direct hits by typhoons, mangrove forests can grow to about 35 m in height. In more exposed areas, however, and in areas with shallower or hard mud, mangroves are reduced to only a few meters in height. In areas with higher salinity (drier climates and less freshwater inflow), mangroves are also shorter and may have saline lagoons behind them instead of more productive back mangroves.

Mangroves in Asia are considered to be the most diverse in the world and covered about 7.7 million ha at the time of the *Global Biodiversity* census (Groombridge, 1992). As shown in Table X, most of this area was in Indonesia (about half in Sumatra), Malaysia (about half in Sabah), Bangladesh, the Philippines, Vietnam, India, Thailand, and Pakistan (see complete statistics in Groombridge, 1992, p. 325). Another 3 million ha belonged to Indonesia but are in Irian Jaya (New Guinea), which geologically is part of Australia, not Asia. These statistics are old, and mangroves have been declining steadily and often rapidly, being converted for development, urban sprawl, or mining (tin); cut for wood; or simply degraded by overuse.

On higher-energy tropical and subtropical coastlines, where the substrate is usually sand, the first beach pioneer inland from the water line is usually the creeping vine *Ipomoea pes-caprae* (Convolvulaceae). Behind this the first stable vegetation is often a dense thicket of large-sclerophyll shrubs, *Scaevola* (Goodeniaceae) and sometimes *Messerschmidtia* (Boraginaceae), plus some pandans and various smaller species which elsewhere might pass as weeds. On many sandy tropical shorelines, coconut palms (*Cocos nucifera*) have established naturally or been planted and cultivated. Strand forests may develop behind beaches and on raised coral terraces, dominated by tree species such as *Pongamia pinnata*, *Ficus microcarpa*, and *Bischoffia javanica*.

At higher latitudes, where winter cold precludes tropical taxa and the previously mentioned systems, salt marshes are the most common shoreline ecosystem. These are also highly productive ecosystems, rivaling tropical rain forests in some cases, and can cover quite wide areas along more concave shorelines, where the intertidal range is higher and thus reaches farther inland. In east Asia, salt marshes extend as far north as the Russian Far East north of Vladivostok. Of course, estuaries also extend to higher latitudes and have been the basis for human settlement in coastal areas for millennia because of their abundance of fish and shellfish. All of these coastline ecosystems, though narrow, are important as buffers between the ocean and the more widespread, typical forests behind them.

Inland, the western Siberian lowland (between the Urals and the Yenisey River, about one-third of all of Siberia) is mostly a vast boreal swamp and may rank as the world's largest single wetland. Boreal landscapes in general, however, are often slowly shifting mosaics of forested uplands and lower lying fens and bogs. In the cold climate most bogs accumulate peat, which raises the land surface such that former depressions become convex surfaces fed only by nutrient-poor precipitation rather than groundwater. These are the highly acidic "raised bogs" which contain mainly ericaceous dwarf-shrubs (e.g., *Oxycoccus* and *Andromeda*), sedges (e.g., *Eriophorum*), and mosses, especially *Sphagnum* species. Around the edges of these (and other) bogs may be stunted birches, larches, and other trees tolerant of saturated soils and high acidity. Although individu-

TABLE X

Mangrove Areas in Asia[a]

Country	Total mangrove area (ha)	Number of protected mangrove areas
East Asia		
China	67,000	24
Hong Kong	0	9
Japan (Ryukyu Islands)	400	4
Taiwan	174	4
Southeast Asia		
Brunei	7,000	3
Cambodia	10,000	0
Indonesia (including Irian Jaya)	4,251,011	152
Malaysia	630,000	99
Myanmar	517,000	6
Philippines	400,000	59
Singapore	1,800	2
Thailand	268,693	17
Vietnam (South)	370,000	2
South and Southwest Asia		
Andaman Islands	(50,000)	25
Bangladesh	410,000	5
India	306,000	9
Iran	23,717	3
Maldives	—	0
Pakistan	249,500	2
Sri Lanka	120,000	9
Arabian Peninsula		
Bahrain	40	1
Saudi Arabia, Oman, Qatar	—	0
United Arab Emirates	—	6
Yemen	—	5

[a] Data are from Groombridge (1992, p. 325).

ally small and not stable over long time scales, these bogs cover a very large area and store much carbon, which may be released to the atmosphere as global warming warms and dries these areas.

Bogs and (less acidic) fens also occur in the temperate zone to the south but especially in moist, cooler areas such as northern Japan. More important in drier climates are probably the freshwater marshes, which are of great importance to birds migrating between wintering areas in tropical Asia (e.g., India) and summer breeding grounds in Siberia. Such marshes may occur in depressions as well as along streams, especially in the deserts of western China, the grasslands of northern China and Mongolia, and in Middle Asia and the highlands of Tibet. These marshes are composed primarily of reeds (*Phragmites*), cattails (*Typha*), sedges (*Cyperus*, *Scirpus*, and *Carex*), and similar herbaceous plants.

Among lakes and streams, the important ecological distinction is between more aerated, nutrient-rich flowing waters and standing waters which can become quite oligotrophic, especially in cooler climates. Lake Baikal, in southern Siberia, is the world's deepest lake (1620 m) and contains about 20% of all the world's (unfrozen) fresh water. The lake contains about 600 plant species (35% endemic) and 1200 animal species (60% endemic). It is also large enough to produce a distinct limnoclimate around its edges, in which species from boreal and polar areas occur in small disjunct populations. Another well-known lake in Asia is Dal Lake in Kashmir, which has a flat bottom, is only a few meters deep, and has been filling rapidly over the past decades with sediment from the surrounding mountains, which have been denuded largely since the 1950s. This and many other lakes are also extremely important for migrating birds. In temperate and tropical Asia, the large rivers are especially important, even though their ecosystems are often badly degraded or at least greatly altered by introductions of exotic species.

TABLE XI

Most Seriously Threatened Wetlands in Asia[a]

Bangladesh
- Chalan Beel*
- Haor Basin of Sylhet and Eastern Mymensingh
- Dubriar Haor*
- The Sundarbans
- Wetlands in Pablakhali Wildlife Sanctuary
- Chokoria Sundarbans*

Bhutan
- Boomthang Valley

Burma (Myanmar)
- Irrawaddy Delta

China
- Yancheng Marshes
- Shijiu Hu
- Shengjin Hu and the lower Yangtze Lakes
- Shengjin Hu
- Xi Jiang (Pearl River) Delta*
- Tuosu Hu (Kurlyk Nor) and Kuerhleiko Hu

India
- Dal Lake
- Wular Lake
- Harike Lake
- Jheels in the vicinity of Haidergarh*
- Dahar and Sauj (Soj) Jheels
- Southern Gulf of Kutch
- Gulf of Khambhat
- Khabartal
- Dipor (Deepar) Bheel
- Logtak Lake
- Salt Lakes Swamp*
- The Sunderbans
- Chilka Lake
- Kolleru Lake
- Estuaries of the Karnataka coast
- Kaliveli Tank and Yedayanthittu estuary
- The Cochin backwaters
- Wetlands in the Andaman and Nicobar Islands

Indonesia
- Banyuasin Musi River Delta
- Muara Cimanuk*
- Sukolilo
- Cilacap and Sagara Anakan
- Danau Bankau, other swamps of Barito Basin*
- Banau Sentarum
- Wetlands in Manusela proposed national park
- Wasur and Rawa Biru

Japan
- Shonai-Fujimae Tidal Rats and Inner Ise Bay
- Lake Shinji and Lake Nakaumi

Korea
- South Kanghwa and North Yongjong mudflats
- Mudflats of South Yongjong and adjacent islands
- Namyang Bay
- Asan Bay
- Kum, Mankyung and Tangjin estuaries

Malaysia
- Sedili Kecil swamp forest
- Klang Islands: Pulau Ketam*
- Kapar Forest Reserve
- North Selangor swamp forest
- Marintaman Mengalong*
- Tempasuk Plain
- Lawas mangroves
- Trusan-Sundar mangroves
- Limbang mangroves
- Maludam swamp forest
- Sarawak Mangrove Forest Reserve

Nepal
- Begnas Tal*

Pakistan
- Khabbaki Lake*
- Siranda Lake*
- Hawkes Bay/Sandspit Beaches and adjacent creeks
- Clifton Beach
- Korangi and Gharo Creeks
- The Outer Indus Delta

Philippines
- Pangasinan Wetlands*
- Manila Bay*
- Laguna de Bay
- Tayabas Bay including Pagbilao Bay
- Inabanga Coast
- Mactan, Kalawisan and Cansaga Bays*
- Agusan marsh
- Lake Leonard*
- Davao Gulf
- Liguasan marsh

Singapore
- Serangoon estuary*

Sri Lanka
- Thandamannar Lagoon*
- Chundikkulam Lagoon
- Chalai Lagoon*
- Periyakarachchi and Sinnakarachchi Lagoons*
- Mahaweli Ganga floodplain system
- Maha Lewaya and Karagan Lewaya
- Lunama Kalapuwa and Kalametiya Kalapuwa*
- Bellanwilla-Attidiya marshes

Taiwan
- Tatu estuary
- Tungshih (Ton-Shou) mangroves*

Thailand
- Gulf of Thailand
- Pak Phanang estuary
- Pa Phru

Vietnam
- Red River Delta
- Red River estuary
- Mekong Delta
- Nam Can mangrove forest

[a] Source: Scott and Poole (1989), via Groombridge (1992, p. 303).

* Sites are considered to be already too degraded to merit any special conservation effort.

TABLE XII

Original and Remaining Areas of Tropical Moist Forest in Asia (km^2)[a]

		Remaining extent			% Moist forest remaining	
Location	Original extent	Atlas maps[b]	Map date	FAO (1988)[c]	Map data	FAO (1988)
Southeast Asia						
Brunei	5,000	4,692	1988	3,230	94	65
Cambodia	160,000	113,250	1971	71,680	71	45
China/Taiwan	340,000	25,860	1979	125,860	8	—
Indonesia	1,700,000	1,179,140	1985–1989	1,138,950	69	67
Laos	225,000	124,600	1987	78,100	55	35
Malaysia	320,000	200,450	—	209,960	63	66
Peninsular	(130,000)	(69,780)	1986	—	54	—
Sabah	(70,000)	(36,000)	1984	—	51	—
Sarawak	(120,000)	(94,670)	1979	—	79	—
Myanmar	600,000	311,850	1987	313,090	52	52
Philippines	295,000	66,020	1988	95,100	22	32
Singapore	500	20	(1980s)	—	4	—
Thailand	250,000	106,900	1985	83,350	43	33
Vietnam	280,000	56,680	1987	75,700	20	27
South Asia						
Bangladesh	130,000	9,730	1981-86	9,270	7	7
India	910,000	228,330	1986	504,010	25	55
Sri Lanka	26,000	12,260	1988	16,590	47	64

[a] Data are from the *Global Biodiversity* report (Groombridge, 1992, p. 262). Subtotals for subunits are in parentheses.

[b] Rain forests and monsoon forests.

[c] Closed broad-leaved plus coniferous forests.

A listing of the most seriously threatened wetlands n Asia is shown in Table XI. Some sites are denoted as already too degraded to merit any special conservation effort. Many other wetlands are not on this priority list but were or have since become threatened to the point that their biodiversity as well as their benefits to local human societies are being rapidly lost.

IX. ARTIFICIAL ECOSYSTEMS

One other class of ecosystems, especially important in ong-settled Asia, is artificial ecosystems such as those nvolving tree plantations, urban landscapes, rice paddies, cropland, managed ponds, etc. Plantations of native conifers are especially extensive in the mountains of Japan, but much more widespread, especially in tropical Asia, are plantations of exotic eucalyptus, Australian pine (*Casuarina*), and other trees not from any part of Asia. Although these plantations, which mainly provide fuelwood, may take the pressure off more natural, native forests, they also damage the soil such that it no longer supports the normal diversity of soil and understory plants, animals, and microbes, especially decomposer organisms which break down and cycle minerals.

Rice paddies and cropland do support some degree of diversity among species which are commonly associated with certain types of management regimes. These largely weedy species may not be species of stable native ecosystems, but they do form somewhat consistent communities. Ponds and other water bodies managed for fish raising, etc. are also common in warmer parts of Asia. The species diversity is low and involves mainly exotic, weedy species such as *Myriophyllum* or water hyacinth (*Eichhornia crassipes*).

Urban landscapes are extensive and ever-growing in Asia as elsewhere. Urban ecosystems may involve allées and other plantings for landscape amelioration but may also involve seminatural wooded, grassy, and aquatic communities. Some of the methods for building seminatural forests in densely built-up areas are quite well developed, especially in Japan (Miyawaki *et al.*, 1987).

X. CONSERVATION STATUS

The conservation status of natural and even of modified but functioning, quasinatural ecosystems in Asia varies widely. Some vast tracts of boreal forest appear un-

TABLE XIII

Total Numbers of Threatened Plant and Vertebrate Species by Country[a]

Country	Plants	Mammals	Birds	Reptiles	Amphibians	Fish
Mediterranean–Southwest Asia						
Afghanistan	4	13	13	1	1	0
Cyprus	43	1	17	1	0	0
Iran	301	15	20	4	0	2
Iraq	1	9	17	0	0	2
Israel	3	8	15	1	1	0
Jordan	752	5	11	0	0	0
Lebanon	5	4	15	1	0	5
Syria	11	4	15	1	0	0
Turkey	1944	5	18	5	1	5
Arabian Peninsula						
Bahrain	0	1	4	0	0	1
Kuwait	1	5	7	0	0	0
Oman	2	6	8	0	0	2
Qatar	0	0	3	0	0	0
Saudi Arabia	2	9	12	0	0	0
United Arab Emirates	0	4	7	0	0	0
Yemen	134	6	9	0	0	0
Tropical Asia						
Bangladesh	33	15	27	14	0	0
Bhutan	15	15	10	1	0	0
British Indian Ocean Territory	0	0	0	0	0	0
Brunei	40	9	10	3	0	2
Cambodia	11	21	13	6	0	5
India	1336	39	72	17	3	2
Indonesia	70	49	135	13	0	29
Laos	3	23	18	5	0	5
Malaysia	522	23	35	12	0	6
Maldives	0	1	1	0	0	0
Myanmar	—	23	42	10	0	2
Nepal	33	22	20	9	0	0
Pakistan	14	15	25	6	0	0
Philippines	159	12	39	6	0	21
Singapore	19	4	5	1	0	1
Sri Lanka	220	7	8	3	0	12
Thailand	68	26	34	9	0	13
Vietnam	338	28	34	8	1	4
East Asia						
China	350	40	83	7	1	7
Hong Kong	5	1	9	2	0	0
Japan	41	5	31	0	1	3
Korea (North)	0	5	25	0	0	0
Korea (South)	33	6	22	0	0	0
Mongolia	0	9	13	0	0	0
Taiwan	95	4	16	0	0	0

[a] Data are from the *Global Biodiversity* report (Groombridge, 1992, p. 239).

touched, but one cannot fly over these areas without seeing smoke (often industrial) rising from what appear to be very isolated places. Other forest areas have almost completely disappeared. It has been estimated that less than 1% of the natural evergreen broad-leaved forest area of Japan remains, and natural broad-leaved as well as conifer forests in China are confined entirely to mountain areas, where they now are also facing encroachment. Tropical forests have also largely disappeared, remaining over large areas mainly in Southeast

Asia (see Table VII). The especially well-developed tropical rain forests of Borneo, with canopy heights commonly near 50 m, have largely been cut within the past 20 years, mostly for export to Japan. Some estimates of original and remaining area of tropical rain forest are shown in Table XII, but the data for remaining area (all pre-1990) must be overestimates.

Country totals of threatened plant and vertebrate species, and also fish, in Asia are summarized in Table XIII. Although these statistics reflect different levels of coverage and different methodologies, they give an indication of the magnitude of the problem of threats to biodiversity by changing conditions—mostly changes in land use by humans. Large numbers of threatened plants are recognized only in Turkey in the west, in Yemen on the Arabian Peninsula, in China in east Asia, and in India, Malaysia, Vietnam, and Sri Lanka in tropical Asia—although many others must also be threatened in other areas. The pattern for animals is more even across countries and regions, but with Indonesia or China recognizing the largest numbers of threatened species in all five categories.

Species and ecosystems are endangered throughout Asia, as elsewhere. For further reading, see in particular the various recent summaries of the status of conservation efforts and strategies in all parts of the world (Hoyt, 1994; Heywood, 1995) as well as more standard treatments of the general problems of wildlife conservation.

See Also the Following Articles

AFRICA, ECOSYSTEMS OF • EUROPE, ECOSYSTEMS OF • NEAR EAST ECOSYSTEMS • NORTH AMERICA, PATTERNS OF BIODIVERSITY IN • SOUTH AMERICA, ECOSYSTEMS OF

Bibliography

Archibold, O. W. (1995). *Ecology of World Vegetation*. Chapman & Hall, London.

Bailey, R. G. (1996). *Ecosystem Geography*. Springer, New York.

Box, E. O. (Ed.), with Peet, R. K., Masuzawa, T., Yamada, I., Fujiwara, K., and Maycock, P. F. (Co-Eds.) (1995). *Vegetation Science in Forestry: Global Perspective Based on Forest Ecosystems of East and Southeast Asia,* Handbook of Vegetation Science, Vol. 12. Kluwer, Dordrecht.

Champion, H. G., and Seth, S. K. (1968). *A Revised Survey of the Forest Types of India*. Government Printer, Delhi.

Groombridge, B. (Ed.) (1992). *Global Biodiversity: Status of the Earth's Living Resources*. World Conservation Monitoring Center/Chapman & Hall, London.

Heywood, V. H. (Ed.) (1995). *Global Biodiversity Assessment*. UN Environment Programme/Cambridge Univ. Press, Cambridge, UK.

Hoyt, J. A. (1994). *Animals in Peril: How "Sustainable Use" Is Wiping Out the World's Wildlife*. Avery, Garden City, NY.

Mani, M. S. (Ed.) (1974). *Ecology and Biogeography in India*. Monographiae Biologicae, Vol. 23. Junk, The Hague.

Missouri Botanical Garden (1983). Biogeographical relationships between temperate eastern Asia and temperate eastern North America. *Ann. Missouri Bot. Garden* **70**, **421–749**.

Miyawaki, A., Bogenrieder, A., Okuda, Sh., and White, J. (1987). *Vegetation Ecology and Creation of New Environments*, Proceedings of 1984 IAVS Symposium. Tokai Univ. Press, Tokyo.

Numata, M. (Ed.) (1974). *The Flora and Vegetation of Japan*. Kodansha/Elsevier, Tokyo/Amsterdam.

Ohsawa, M. (1995). Latitudinal comparison of altitudinal changes in forest structure, leaf type, and species richness in humid monsoon Asia. *Vegetatio* 121, 310.

Raven, P. H., and Axelrod, D. I. (1974). Angiosperm biogeography and past continental movements. *Ann. Missouri Bot. Garden* **61**, **539–673**.

Schaller, G. B. (1998). *Wildlife of the Tibetan Steppe*. Univ. of Chicago Press, Chicago.

Scott, D. A., and Poole, C. M. (1989). *A Status Overview of Asian Wetlands,* No. 53. AWB, Kuala Lampur, Malaysia.

Troll, C. (Ed.) (1972). *Geoecology of the High Mountain Regions of Eurasia*, Erdwissenschaftl. Forschung, Vol. 4. Franz-Steiner-Verlag, Wiesbaden.

Udvardy, M. D. F. (1975). IUCN Occasional Paper No. 18. International Union for the Conservation of Nature and Natural Resources, Morges, Switzerland.

Walter, H. (1974). *Die Vegetation Osteuropas, Nord- und Zentralasiens*, "Vegetation der einzelnen Groβräume" series, Vol. VII. Gustav-Fischer-Verlag, Stuttgart.

Walter, H. (1985). *Vegetation of the Earth and Ecological Systems of the Geobiosphere*, 3rd ed. Spinger-Verlag, New York.

Walter, H., and Box, E. O. (1983). Deserts of Eurasia. In *Temperate Deserts and Semi-Deserts* (N. E. West, Ed.), Ecosystems of the World, Vol. 5. Elsevier, Amsterdam.

Wang, C. W. (1961). *The Forests of China, with a Survey of Grassland and Desert Vegetation*, Maria Moors Cabot Foundation Publ. No. 5. Harvard Univ., Cambridge, MA.

Whitmore, T. C. (1984). *Tropical Rainforests of the Far East*, 2nd ed. Clarendon, Oxford, UK. (1st ed., 1975)

Wu, Zhengyi, and Wu, Sugong (1996). In *Floristic Characteristics and Diversity of East Asian Plants* (A.-L. Zhang and S.-G. Wu, Eds.). Springer-Verlag, Berlin.

Wu, Zhengyi, and Wu, Sugong (1998). A proposal for a new floristic kingdom (realm)—the East Asiatic Kingdom, its delineation and characteristics. *In:* Floristic Characteristics and Diversity of East Asian Plants (Zhong, A., and Wu, S., eds.), pp. 3–42. China Higher Education Press, Bejing, and Springer-Verlag, Berlin.

Zohary, M. (1973). *Geobotanical Foundations of the Middle East*, "Geobotanica Selecta" series, Vol. 3. Gustav-Fischer-Verlag, Stuttgart.

ATMOSPHERIC GASES

Donald J. Wuebbles
University of Illinois at Urbana-Champaign

GLOSSARY

atmospheric lifetime Timescale characterizing the rate of removal of a gas from the atmosphere, generally defined as the time to remove 63.2%, the e-folding time.

greenhouse gases Gases like water vapor and carbon dioxide that can affect climate through their absorption and reemission of terrestrially emitted infrared radiation.

mixing ratio Ratio of the concentration of a gas to the concentration of air. It can be defined in terms of volume (e.g., molecules per m^3) or mass (e.g., g per m^3).

oxidizing capacity Self-cleansing ability of the atmosphere through oxidative reactions.

ozone Molecule composed of three oxygen atoms that is extremely important to life on Earth because of its absorption of solar ultraviolet radiation; ozone is also a greenhouse gas.

photochemical source or sink Production or loss of an atmospheric gas involving photodissociation and/or chemical reactions.

photodissociation Destruction of a molecule through absorption of solar radiation and subsequent breaking of one or more chemical bonds.

THE EARTH'S ATMOSPHERE IS COMPOSED OF A MIXTURE OF GASES THAT WE CALL AIR. The most abundant of these gases, molecular nitrogen and molecular oxygen, are not being greatly affected by human activities and are unlikely to be so in the foreseeable future. However, many of the other gases, like carbon dioxide and ozone, are being affected by human activities. Although there are only trace amounts of these gases in the atmosphere, many of them are extremely important to life on Earth and to biodiversity. This article describes many of the important atmospheric gases, their role in affecting the planet, and the factors affecting their changing concentrations in the atmosphere.

I. INTRODUCTION

As seen in Table I, the bulk of the Earth's dry atmosphere is composed of molecular nitrogen (N_2, 78.08%), molecular oxygen (O_2, 20.95%), and argon (Ar, 0.93%). While the remaining gases comprise only about 0.03% of the Earth's atmosphere, many of them are extremely important to the climate and habitability of our planet. These gases are the primary focus of this article. The concentrations of many of these gases are also being affected by human activities.

TABLE I

Typical Concentrations and Primary Sources of Some of the Important Gases in the Atmosphere

Constituent	Chemical formula	Volume mixing ratio (dry air)[a]	Major sources and remarks
Nitrogen	N_2	78.084%	Biological
Oxygen	O_2	20.948%	Biological
Argon	Ar	0.934%	Inert
Carbon dioxide	CO_2	360 ppmv	Combustion, ocean, biosphere
Methane	CH_4	1.7 ppmv	Biogenic, human-related
Hydrogen	H_2	0.55 ppmv	Biogenic, human-related, photochemical
Nitrous oxide	N_2O	0.31 ppmv	Biogenic, human-related
Carbon monoxide	CO	0.05–0.2 ppmv	Photochemical, human-related
Ozone (troposphere)	O_3	0.01–0.5 ppmv	Photochemical
Ozone (stratosphere)	O_3	0.5–10 ppmv	Photochemical
Nonmethane hydrocarbons	C_2H_2, etc.	5–20 ppbv	Biogenic, human-related
CFC-11	$CFCl_3$	260 pptv	Human-related
CFC-12	CF_2Cl_2	530 pptv	Human-related
CFC-113	CCl_2FCClF_2	80 pptv	Human-related
Halon-1211	$CBrClF_2$	3.5 pptv	Human-related
Halon-1301	$CBrF_3$	2.3 pptv	Human-related
Methyl bromide	CH_3Br	10 pptv	Ocean, human-related
Methyl chloride	CH_3Cl	550 pptv	Ocean, biogenic, human-related
Nitrogen species	NO_y	10 pptv–1 ppmv	Soils, lightning, human-related
Ammonia	NH_3	10 pptv–1 ppbv	Biogenic
Hydroxyl	OH	0.1 pptv–10 pptv	Photochemical
Peroxyl	HO_2	0.1 pptv–10 pptv	Photochemical
Hydrogen peroxide	H_2O_2	0.1–10 ppbv	Photochemical
Formaldehyde	CH_2O	0.1–1 ppbv	Photochemical
Sulfur dioxide	SO_2	10 pptv–1 ppbv	Photochemical, volcanic, human-related
Dimethyl sulfide	CH_3SCH_3	10–100 pptv	Biogenic
Carbon disulfide	CS_2	1–300 pptv	Biogenic, human-related
Carbonyl sulfide	OCS	500 pptv	Biogenic, volcanic, human-related
Hydrogen sulfide	H_2S	5–500 pptv	Biogenic, volcanic

[a] Ppmv = parts per million by volume = number of molecules in 10^6 molecules of air; ppbv = parts per billion by volume = number of molecules in 10^9 molecules of air; pptv = parts per trillion by volume = number of molecules in 10^{12} molecules of air.

Some of the gases listed in Table I, like ozone (O_3), affect the transmission of solar radiation in the atmosphere. Others, like carbon dioxide (CO_2), affect the absorption of terrestrially produced infrared radiation and are called greenhouse gases. These types of gases are therefore linked to the physical climate system of the Earth. Increasing concentrations of greenhouse gases have resulted in concerns about global warming. Some trace gases, such as sulfur dioxide (SO_2) and hydrogen sulfide (H_2S), can affect climate through their role as precursors to the production of atmospheric particles. The resulting sulfate particles can scatter solar radiation, thus preventing some of it from reaching the Earth's surface. These particles can also serve as condensation nuclei in the formation of clouds.

The ozone layer is a term that refers to the distribution of ozone that is naturally formed in the stratosphere. This layer protects life on Earth from harmful levels of solar ultraviolet radiation. Chlorofluorocarbons (CFCs) and other chlorinated and brominated halocarbons are emitted by a variety of human activities. Atmospheric measurements have clearly corroborated theoretical studies showing that the chlorine and bromine released from the destruction of these halocarbons in the stratosphere have been reacting to destroy significant amounts of ozone over the last few decades.

Gases like hydroxyl (OH), ozone, and hydrogen peroxide (H_2O_2) control the oxidizing capacity of the atmosphere, and destroy many pollutants emitted into the atmosphere through chemical reactions. Hydroxyl has been referred to as the "atmospheric vacuum cleaner." It is generated primarily by the interactions of water vapor, ozone, and solar ultraviolet radiation, and is destroyed primarily through reaction with carbon monoxide (CO) and methane (CH_4).

Emissions of trace gases in urban regions can result in atmospheric pollution referred to as photochemical smog. For example, ozone, a potential cause of health problems in polluted regions, can be formed in the air by chemical reactions following emissions of nitrogen oxides (NO_x) and volatile organic gases (VOCs), including hydrocarbons. NO_x is the sum of the concentrations of nitric oxide (NO) and nitrogen dioxide (NO_2), while NO_y also includes other reactive nitrogen gases like nitric acid (HNO_3).

Missing from Table I is water vapor, a gas with highly variable concentration, whose amount in the atmosphere is dependent on evaporation and hydrological cycle processes associated with weather. The concentration of atmospheric water vapor can be affected at the local scale by human activities but is not thought to be greatly affected at the global scale by human activities.

Biodiversity affects the concentrations of atmospheric trace gases and can also be affected by changing concentrations of atmospheric gases. The biosphere is an important component in the biogeochemical cycles of carbon, nitrogen, and sulfur. As a result, the sources and sinks of many atmospheric gases are strongly affected by changes in the biosphere. For example, deforestation is contributing to the increase in concentration of carbon dioxide. Local air pollution, changes in climate, and changes in stratospheric ozone can affect the biosphere and have an impact on biodiversity. For example, acid rain or acidification resulting from emissions of nitrogen- and sulfur-containing gases can affect landscapes in many ways.

II. THE CHANGING COMPOSITION

Without human intervention, concentrations of many atmospheric gases would be expected to change slowly. Ice core measurements of the gases trapped in ancient ice bubbles indicate that this was the case before the last century. However, since the beginning of the Industrial Age, emissions associated with human activities have risen rapidly. Agriculture, industry, waste disposal, deforestation, and especially fossil fuel use have been producing increasing amounts of carbon dioxide, methane, nitrous oxide (N_2O), chlorofluorocarbons, and other important gases. Because of these increasing emissions, atmospheric levels of these gases have been building at an unprecedented rate, resulting in effects on ozone, another radiatively important gas, and raising concerns regarding the impact of these gases on climate. Although all of these gases are having some influence on climate, carbon dioxide is of particular interest in the concerns about global warming. In addition, changes in the human-related emissions and atmospheric concentrations of small particles or aerosols are also influencing climate. The following discussion focuses on the changing concentrations and budgets of a number of these gases.

A. Carbon Dioxide

Of all the greenhouse gases, carbon dioxide has been undergoing the largest changes in concentration. It is also the gas of most concern to analyses of the potential human effects on climate. Accurate measurements of atmospheric CO_2 concentration began in 1958 at the Mauna Loa Observatory in Hawaii. Figure 1 shows that the annually averaged concentration of CO_2 in the atmosphere has risen from 316 ppm (parts per million) in 1959 to 364 ppm in 1997. The CO_2 measurements exhibit a seasonal cycle, which is mainly caused by the seasonal uptake and release of atmospheric CO_2 by terrestrial ecosystems. The average annual rate of

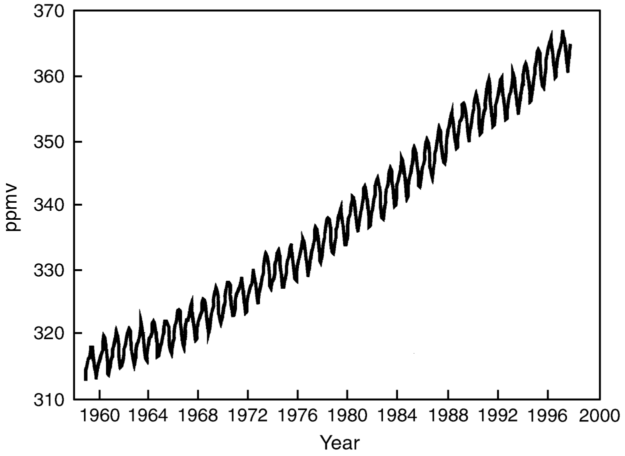


FIGURE 1 Observed monthly average of carbon dioxide (CO_2) concentration (ppmv, parts per million by volume) from Mauna Loa, Hawaii (measurements by C. Keeling and colleagues). Seasonal variations are primarily due to the uptake and production of CO_2 by the terrestrial biosphere.

increase over the whole time period is about 1.2 ppmv or 0.4% per year, with the rate of increase over the last decade being about 1.6 ppmv/yr. A significant slowing in the growth rate occurred in 1992 (when it dropped to 0.6 ppmv), but by 1994 the rate had risen above the average over the decade. This anomaly in the carbon cycle system appears to be related to atmospheric aerosols resulting from the Mount Pinatubo volcanic eruption in June, 1991 and the corresponding effects on solar irradiances and natural emissions and/uptake of CO_2.

Measurements of CO_2 concentration in air trapped in ice cores indicate that the pre-industrial concentration of CO_2 was approximately 280 ppmv. These data show that carbon dioxide concentrations fluctuated by ±10 ppmv around 280 ppmv for over a thousand years until the recent increase to the current 360+ ppmv, an increase of over 30%.

Why has the atmospheric concentration of CO_2 increased so dramatically? Analyses with models of the atmosphere–ocean–biosphere system of the carbon cycle indicate that human activities are primarily responsible for the increase in CO_2. Two types of human activities are primarily responsible for emissions of CO_2: fossil fuel use, which released about 6.0 GtC (gigatons of carbon; a gigaton is 10^9 tons) into the atmosphere in 1990, and land use, including deforestation and biomass burning, which may have contributed about 1.6 ± 1.0 GtC in addition to that from fossil fuels.

This added atmospheric carbon resulting from human activities is redistributed within the atmospheric, oceanic, and biospheric parts of the global carbon cycle, with the dynamics of this redistribution determining the corresponding rise in atmospheric CO_2 concentration. Atmospheric CO_2 increased at the rate of 3.4 ± 0.2 GtC/yr during the 1980s. Recent analyses of oceanic and atmospheric data indicate that the likely rate of oceanic uptake is 2.0 ± 0.8 GtC/yr. In the future, as the amount of CO_2 increases in the atmosphere and in the ocean, it is expected that the oceans will take up a smaller percentage of the new emissions.

Analyses of the carbon budget have implied that there is a mismatch between observed levels of CO_2 and known loss processes. This discrepancy suggests that a missing carbon sink has existed during recent decades. This sink now appears to be explainable in terms of increased net carbon storage by the terrestrial biomass stimulated by the CO_2 fertilization effect (increased plant growth in an atmosphere of higher CO_2 concentration) and other processes. Based on the calculated ocean uptake rate, together with estimated fossil emissions and the observed atmospheric inventory change, the terrestrial biosphere uptake in the 1980s was 1.3 ± 1.5 GtC/yr.

1. Emissions from Fossil Fuel Consumption

Carbon dioxide is emitted when carbon-containing fossil fuels are oxidized by combustion. Carbon dioxide emissions depend on energy and carbon content, which ranges from 13.6 to 14.0 MtC/EJ for natural gas, 19.0 to 20.3 for oil, and 23.9 to 24.5 for coal. Other energy sources, such as hydro, nuclear, wind, and solar, have no direct carbon emissions. Biomass energy, however, is a special case. When biomass is used as a fuel, it releases carbon with a carbon-to-energy ratio similar to that of coal. However, the biomass has already absorbed an equal amount of carbon from the atmosphere prior to its emission, so that net emissions of carbon from biomass fuels are zero over its life cycle.

Human-related emissions from fossil fuel use have been estimated as far back as 1751. Before 1863, emissions did not exceed 0.1 GtC/yr. However, by 1995 they had reached 6.5 GtC/yr, giving an average emission growth rate slightly greater than 3% per year over the last two and a half centuries. Recent growth rates have been significantly lower, at 1.8% per year between 1970 and 1995. Emissions were initially dominated by coal. Since 1985, petroleum products have been the main source of emissions despite their lower carbon content. The regional pattern of emissions has also changed. Global emissions were once dominated by Europe and North America, but now developing nations are providing an increasing share of emissions. In 1995, developing countries (including China and India as part of the Annex I nations being used in international policy discussions) accounted for 48% of global emissions.

2. Emissions from Land Use Changes

The biosphere holds approximately 560 GtC in the form of aboveground biomass, and an additional 1200 GtC in soils and detritus. These pools form the principal reservoirs from which terrestrial systems can exhaust or sequester carbon. Evaluations of carbon releases from vegetation and soils based on changes in land use indicate that, since 1800, land use practices have decreased carbon storage in vegetation and soil by about 170 Gt. Estimates of the net flux of carbon into the atmosphere in 1980 due to land use range from 0.4 to 2.5 GtC/yr. For the decade of the 1980s, tropical emissions from changes in land use averaged 1.6 ± 1.0 GtC/yr.

B. Methane

Although methane's atmospheric abundance is less than 0.5% that of CO_2 on a molecule-by-molecule basis, a molecule of CH_4 is approximately 50 times more effective as a greenhouse gas in the current atmosphere than CO_2. When this is combined with the large increase in its atmospheric concentration, methane becomes the second most important greenhouse gas of concern to climate change. Based on analyses of ice cores, the concentration of methane has more than doubled since preindustrial times. The current globally averaged atmospheric concentration of methane is about 1.73 ppmv (Fig. 2). Continuous monitoring of methane trends in ambient air from 1979 to 1989 indicates that concentrations have been increasing at an average of about 16 ppbv (parts per billion by volume) or 1% per year.

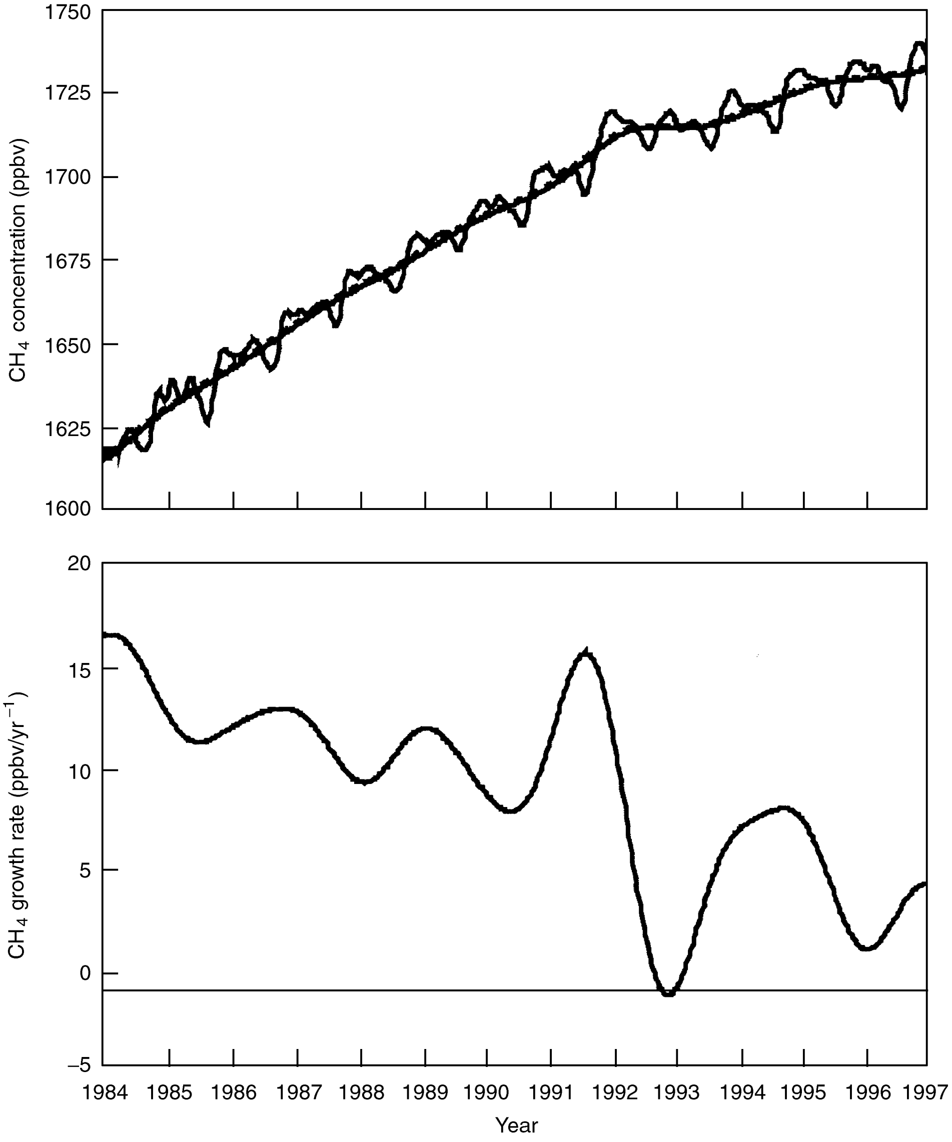


FIGURE 2 (Top) Globally averaged atmospheric methane (CH_4) concentrations (ppbv, parts per billion by volume) derived from the NOAA Climate Monitoring and Diagnostics Laboratory air-sampling sites by E. Dlugokencky and colleagues. The solid line is a deseasonalized trend curve fitted to the data. The dashed-line is a model-estimated (that accounts for methane emissions and loss in the atmosphere) calculated trend that fit to the globally average values. (Bottom) Atmospheric CH_4 instantaneous growth rate (ppbv/yr), which is the derivative with respect to the trend curve shown in the first panel.

In the late 1980s, rates of increase fell to about 10 ppbv/yr; they declined even further in the early 1990s before stopping entirely at some locations in 1992. In 1994, global methane growth rates recovered to about 8 ppbv/yr. There are a number of ideas to explain these rapid changes in the growth rate, ranging from suggested reductions in emissions from anthropogenic or natural sources to a slowing in the rate of CH_4 removal. However, the best indications are that the sharp increase and the subsequent dip in the early 1990s are connected to changes in atmospheric chemistry and temperature induced by the 1991 Pinatubo eruption. Yet the cause of the longer-term global decline in methane growth is still not well understood, although it may be that much of the earlier rapid increase in methane emissions from agricultural sources is now slowing down. A recent increase in the trend of methane during the last one and a half years (not shown in Fig. 2) to ~10 ppbv/yr is also not fully understood but may be associated with the response of wetlands to increasing global temperatures over the last several years.

1. The Methane Budget

Methane emissions come from a number of different sources, both natural and anthropogenic. One type of human-related emissions arises from biogenic sources from agriculture and waste disposal, including enteric fermentation, animal and human wastes, rice paddies, biomass burning, and landfills. Emissions also result from fossil fuel-related methane sources such as natural gas loss, coal mining, and the petroleum industry. Methane is emitted naturally by wetlands, termites, other wild ruminants, oceans, and hydrates. Based on recent estimates, current human-related biogenic and fossil fuel-related sources for methane are approximately 275 and 100 Tg (teragrams, or 10^{12} g) CH_4/yr, respectively, while total natural sources are around 160 $TgCH_4$/yr.

Because of the variety of methane sources, emissions are affected by numerous factors, including energy use, human population distributions, agricultural practices, and climate. These factors complicate the resolution of past emissions and make predictions of future CH_4 emissions difficult. Estimates of historical methane emissions can be obtained by analyses of past measurements of atmospheric CH_4 concentrations and by analyses of emissions from the different sources. Such studies suggest that, although methane emissions have climbed rapidly over the past four decades, the relative importance of agricultural sources may be declining, whereas the importance of nonagricultural sources related to fossil fuel use and waste disposal is on the increase.

In contrast to the numerous sources of methane, there are only one major and two minor sinks for tropospheric methane. Reaction with the hydroxyl radical is responsible for the removal of approximately 490 $TgCH_4$/yr (88% of the total sink). The remainder of the CH_4 is removed through reactions with soil (30 $TgCH_4$/yr, or ~5%) or transport to the stratosphere (40 $TgCH_4$/yr, or ~7%).

C. Nitrous Oxide

Nitrous oxide is a greenhouse gas that on a molecule-to-molecule basis is 200 times more efficient than CO_2 in absorbing infrared radiation. Also, through reactions with excited oxygen atoms, N_2O is the primary source of the nitrogen oxides that account for a significant fraction of the natural destruction of ozone in the stratosphere. Atmospheric measurements and ice core data indicate a continuous increase in N_2O from a pre-industrial concentration of about 275 ppbv. In 1990, the mean atmospheric concentration of N_2O was about 311 ppbv, with a current growth rate of 0.2–0.3% per year.

Nitrous oxide is produced by various natural and human-related sources, most of which are not well quantified. Current estimates of mean emissions from natural sources include: oceans, 3 TgN/yr; tropical soils (wet forests, dry savannas), 4 TgN/yr; and temperate soils (forests, grasslands), 2 TgN/yr. Anthropogenic sources include fertilized cultivated soils, 3.5 TgN/yr; biomass burning, 0.5 TgN/yr; industrial sources, 1.3 TgN/yr; and cattle and feedlots, 0.4 TgN/yr. Large uncertainties associated with the numerous small sources that make up the N_2O budget make it difficult to fully explain its increase in concentration.

The major sink for N_2O is photodissociation by sunlight in the stratosphere. The current best estimate for stratospheric removal, based on stratospheric chemistry modeling, is 12.5 TgN/yr. There is some evidence that N_2O is consumed by certain soils, but there are not enough data to make a reasonable global estimate of this sink.

D. Chlorofluorocarbons and Other Halocarbons

Halocarbons are greenhouse gases that can contribute to climate change, and they are also largely responsible for stratospheric ozone loss over recent decades. Because of their dual impact, they are of particular concern to environmental health. Among the most potent halocarbons in the current atmosphere are the chlorofluorocarbons CFC-11 ($CFCl_3$) and CFC-12 (CF_2Cl_2). One molecule of CFC-11 or CFC-12 in the atmosphere is, respectively, 12,400 and 15,800 times more effective as a greenhouse gas than one molecule of CO_2. With the exception of the naturally occurring portions of CH_3Cl and CH_3Br emissions, all the halocarbons in the atmosphere are man-made. Their inertness and long lifetimes have made them attractive chemicals for use as propellants, refrigerants, fire retardants, and other industrial applications. Natural sources of CH_3Cl from ocean surface waters and wood-rotting fungi may account for as much as half of atmospheric CH_3Cl, with biomass burning (some of which is natural) being the other primary source. Roughly half of the CH_3Br comes from the oceans, while biomass burning and its use as a soil fumigant account for the majority of its remaining emissions into the atmosphere.

Halocarbons containing chlorine and/or bromine are of particular concern with regard to destruction of stratospheric ozone. Bromine and chlorine effectively catalyze ozone destruction cycles. These halocarbons, plus others such as the perfluorocarbons (PFCs) and the hydrofluorocarbons (HFCs) that contain fluorine instead of chlorine, also have the potential to affect climate change since these chemical species characteristically have strong infrared absorption features.

Measurements of CFCs and other compounds by the NOAA Climate Monitoring and Diagnostics Laboratory at sites throughout the world are shown in Fig. 3. CFC-11 and CFC-12 have the largest atmospheric concentrations, at 0.26 and 0.53 ppbv, respectively. The tropospheric concentrations of both of these gases were increasing at about 4% per year in the early 1990s, but have now slowed appreciably. The concentrations of CFC-11 and several other controlled halocarbons have already started to decline. The use of these compounds has diminished greatly, with all but essential applications being banned by the beginning of 1996. The atmospheric concentrations of several other halocarbons have, until recently, been growing at an even faster rate than CFC-11 and CFC-12. For example, the concentration of CFC-113 ($C_2F_3Cl_3$) was increasing about 10% per year in the early 1990s but has also slowed greatly, with a current concentration of about 0.08 ppbv. Abundances of the shorter-lived controlled compound methyl chloroform (CH_3CCl_3), have declined appreciably. HCFC-22 (CHF_2Cl), a refrigerant often found in home air conditioners, has gained increased use as a replacement for CFCs and its concentration has been increasing at about 5–6% per year since 1995. The measured concentrations of several other replacements are also increasing.

All of the fully halogenated chlorofluorocarbons have long atmospheric lifetimes. The lifetime of CFC-11 is about 50 years, while the lifetime of CFC-12 is about 102 years. The atmospheric lifetimes of HCFCs (hydrochlorofluorocarbons), HFCs, and other halocarbons containing hydrogen tend to be much shorter than those of the CFCs. Because of these shorter lifetimes, less of the replacement compounds containing chlorine and bromine reach the stratosphere and they have less effect on ozone than the CFCs.

Bromine is more effective at destroying ozone than chlorine. Therefore, early rapid increases in atmospheric concentrations of bromine-containing halons, most notably Halon-1301 (CF_3Br) and Halon-1211 (CF_2ClBr), have caused concern. Despite their control under the 1997 Montreal Protocol on Substances That Deplete the Ozone Layer, the concentrations of several halons, including H-1211 and H-1301, continue to increase. Primary destruction of these compounds occurs through photolysis, resulting in long atmospheric lifetimes (65 years for H-1301 and 20 years for H-1211). However, these compounds currently have small atmospheric concentrations, about 4 pptv (parts per thousand by volume) or less, and hence their contribution to the absorption of infrared radiation is considered minimal.

E. Other Gases

1. Carbon Monoxide

In addition to its potential effects on health in polluted areas, carbon monoxide affects the oxidizing capacity of the atmosphere through its reactivity. The reaction of carbon monoxide with OH is the primary sink for atmospheric OH. This reaction also yields an additional source of the greenhouse gas CO_2.

The global emissions of CO are still poorly understood. Sources of carbon monoxide include incomplete combustion processes (complete combustion yields CO_2 rather than CO). Emissions of CO from fossil fuel combustion peak between 30 and 60 degrees north latitude. More than 70% of the CO from the biomass burning source is emitted in tropical regions. Chemical decomposition of methane and other hydrocarbons is also an important source. It is estimated that about two-thirds of current atmospheric CO results from human activities. Total sources are about 1100 TgC per year with a large uncertainty range. As a result of its short lifetime (2–3 months) and the high spatial variability of its sources, the atmospheric concentration of CO varies greatly in time and space. Annually averaged concentrations of CO peak at about 0.2 ppmv at high northern latitudes; minimum concentrations of about

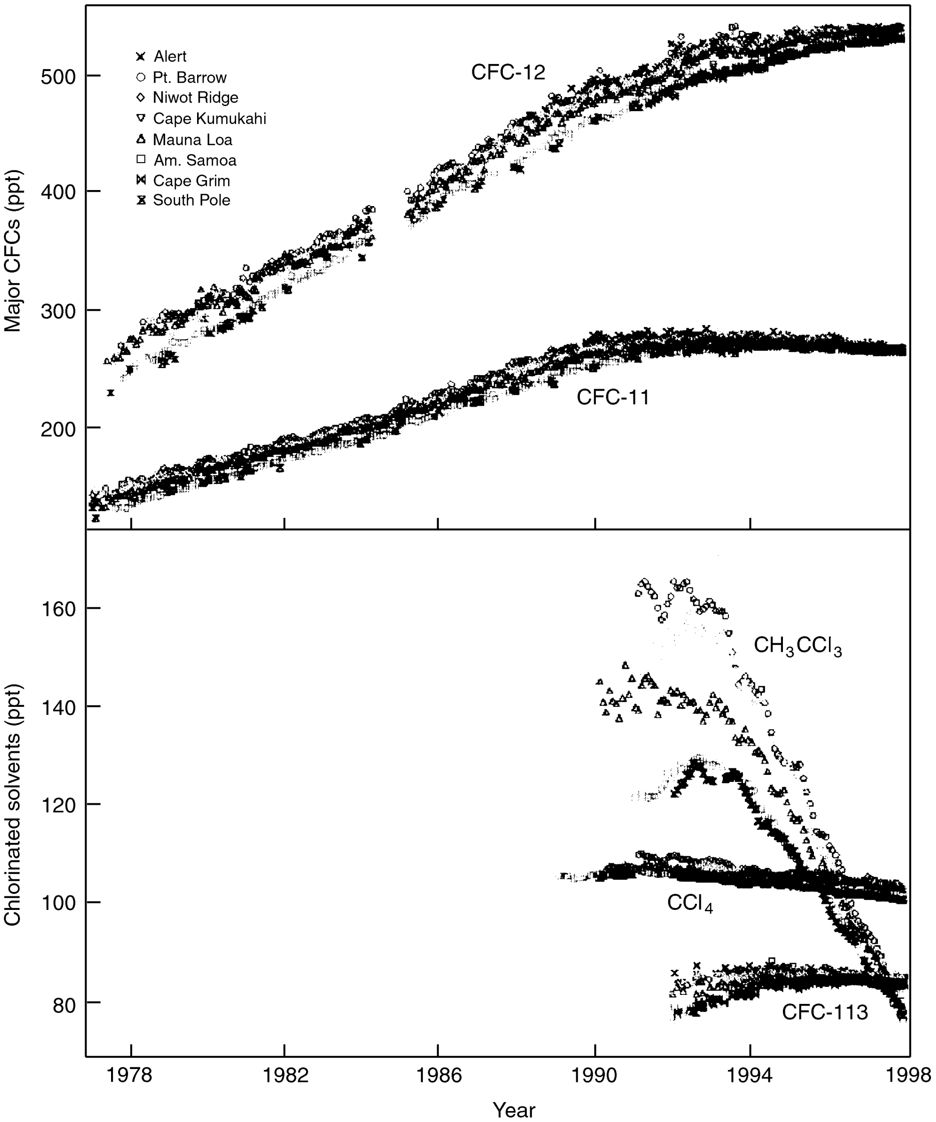


FIGURE 3 (Top) Changes in measured chlorofluorocarbon concentrations over time, showing a sharp rise over the last few decades and the beginning of a recent decline. (Bottom) Changes in concentrations of several major chlorinated solvents over the last decade. Data are based on measurements from the NOAA Climate Monitoring and Diagnostics Laboratory (J. Elkins and colleagues).

0.06 ppmv occur throughout most of the Southern Hemisphere. Long-term trends up to 1990 suggest that CO concentrations had been increasing in the Northern Hemisphere at about 1% per year, with little evidence of an increase in the Southern Hemisphere. After 1990, concentrations appear to have decreased.

2. Non-methane Hydrocarbons

Nonmethane hydrocarbons (NMHCs) in the presence of nitrogen oxides (NO_x) contribute to the formation of tropospheric ozone and stratospheric H_2. They can also react with nitrogen species to produce the atmospheric gas peroxyacetyl nitrate (PAN), which is a long-lived reservoir of reactive nitrogen. Oxidation of NMHCs is an additional source of CO and ultimately CO_2.

There are many anthropogenic and natural sources emitting NMHCs into the atmosphere. Anthropogenic sources result from all aspects of human activity, including chemical manufacturing, vehicle exhaust, food processing, refuse disposal, biomass burning, and energy production. Global emissions from anthropogenic activities are estimated at roughly 140 Tg/yr. Emissions from the natural sources are less well known.

3. Nitrogen Oxides

Emissions of nitrogen oxides have been a cause of concern because of their role as primary pollutants in photochemical smog and their contribution to acid wet and dry deposition. Nitrogen oxides are also important because of their indirect effect on climate through their role in affecting global ozone concentrations. Although NO_x species are relatively short-lived, they can react chemically with hydrocarbons to produce PAN. The constituent PAN provides a reservoir for nitrogen oxides that can be transported long distances to affect ozone chemistry well downstream from the sources. The lifetime of PAN depends strongly on temperature.

There is significant uncertainty in the sources of nitrogen oxides, with the total source being estimated to be roughly 53 Tg (N)/yr. The largest sources of reactive nitrogen in the troposphere are fossil fuel combustion, biomass burning, lightning discharges, microbial activity in soils, aircraft emissions, and transport from the stratosphere. Emissions of nitrogen oxides from combustion of fossil fuel have increased globally at 1–2% per decade during this century, resulting in increased tropospheric concentrations particularly over continents in the lower atmosphere and in the flight corridors used by commercial aircraft. Nitrogen dioxide is an important absorber of visible solar radiation, and it could affect climate directly if tropospheric and/or stratospheric concentrations continue to increase.

4. Sulfur Gases

Emissions of sulfur dioxide and other gases can result in the formation of aerosols that can affect climate. Aerosols affect climate directly by absorption and scattering of solar radiation and indirectly by acting as cloud condensation nuclei (CCN). A variety of analyses indicate that human-related emissions of sulfur, and the resulting increased sulfuric acid concentrations in the troposphere, may be cooling the Northern Hemisphere sufficiently to compensate for much of the warming expected from greenhouse gases. Volcanic emissions can influence climate for short periods (1 to 3 years) through emissions of sulfur dioxide into the lower stratosphere.

Over half of the sulfur dioxide (SO_2) emitted into the atmosphere comes from human-related sources, mainly from the combustion of coal and other fossil fuels. Most of these emissions occur in the Northern Hemisphere. Analyses indicate that anthropogenic emissions have grown dramatically during this century. Other SO_2 sources come from biomass burning, from volcanic eruptions, and from the oxidation of di-methyl sulfide (DMS) and hydrogen sulfide (H_2S) in the atmosphere. DMS and H_2S are primarily produced in the oceans. Atmospheric SO_2 has a lifetime of less than a week, leading to formation of sulfuric acid and eventually to sulfate aerosol particles. Gas-to-particle conversion can also occur in cloud droplets; when precipitation doesn't occur soon, the evaporation of such droplets can then leave sulfate aerosols in the atmosphere.

During periods of low volcanic activity, carbonyl sulfide (COS) is thought to be responsible for the maintenance of the sulfuric aerosol layer found in the lower stratosphere. Natural emissions explain most of the COS in the present atmosphere, while the relatively long atmospheric lifetime (about 2 years) of COS explains why much of it reaches the stratosphere before its conversion to sulfuric acid aerosol. However, if sources of COS (or its precursor, CS_2) were to increase dramatically, the background aerosol layer concentration would increase, with significant implications for climate.

III. THE IMPORTANCE OF STRATOSPHERIC OZONE

Ozone, O_3, is composed of three oxygen atoms and is a gas at atmospheric pressures and temperatures. Most of the ozone (about 90%) exists in the stratosphere, the layer of the atmosphere about 10–50 km above the Earth's surface. The remaining ozone is in the troposphere, the lower region of the atmosphere extending from the Earth's surface up to roughly 10 km at midlatitudes and 16 km in the tropics. Figure 4 shows a typical integrated column of ozone, referred to as the total ozone column, as a function of latitude and season based on the satellite observations from the TOMS (Total Ozone Mapping Spectrophotometer) instrument. Despite the fact that the primary production of ozone occurs in the tropics and the mid-latitudes, the largest amounts of ozone are found at high latitudes as a result of the pole-ward transport of ozone by atmospheric dynamical processes. The large decrease in ozone over springtime Antarctica, termed the Antarctic ozone "hole," can also be seen in Fig. 4.

Ozone in the troposphere and stratosphere is chemically identical, but it has very different effects on life

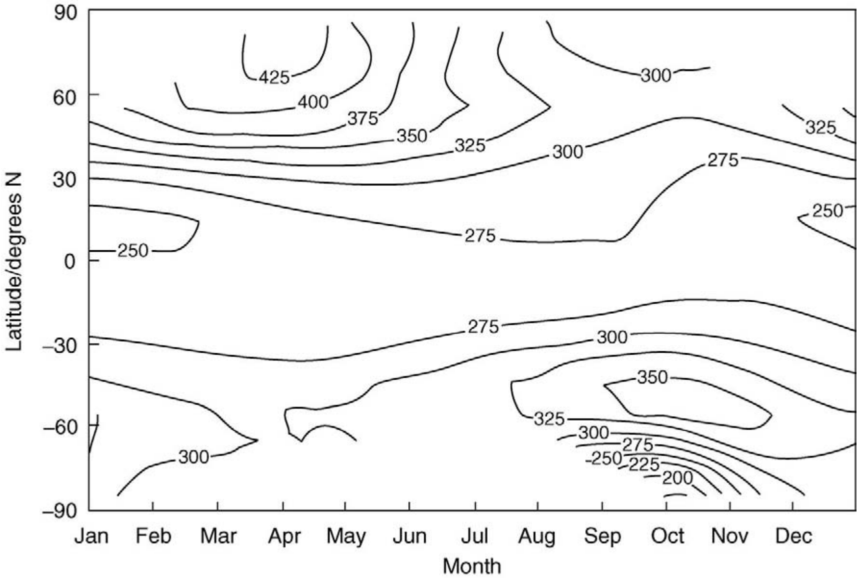


FIGURE 4 Integrated column of ozone, referred to as the total ozone column, as a function of latitude and season based on Nimbus-7 TOMS (Total Ozone Mapping Spectrophotometer) satellite observations.

on the Earth depending on its location. Stratospheric ozone plays a beneficial role by absorbing solar ultraviolet radiation (UV-B), thus preventing biologically harmful levels of UV radiation from reaching the Earth's surface. It is the absorption of solar radiation by ozone that explains the increase in temperature with altitude in the stratosphere. Concerns about increased UV-B from the decreasing levels of ozone have been the driver for policy actions to protect the ozone layer.

Ozone is also a greenhouse gas, with a large infrared absorption band in the atmospheric window, at 9.6 μm. It is the balance between the solar and infrared radiative processes that determines the net effect of ozone on climate. Increases in ozone in the stratosphere above about 30 km tend to decrease the surface temperature as a result of the increased absorption of solar radiation, effectively decreasing the solar energy that would otherwise warm the Earth's surface. Below 30 km, increases in ozone tend to increase the surface temperature, and the infrared greenhouse effect dominates in this region.

Closer to the Earth's surface, ozone displays its destructive side. Ozone is a strong oxidizer. Hence, direct exposure to high levels of ozone has toxic effects on human health and plants. Although ozone is a major component of photochemical smog in urban areas, this ozone is generally not thought to be a significant contributor to the global ozone budget. Balloon measurements suggest that tropospheric ozone at the global scale has been increasing.

A. Stratospheric Ozone Trends

Concentrations of ozone in the stratosphere result from chemical production and destruction processes in combination with transport processes. Production of ozone in the stratosphere results primarily from photodissociation of oxygen molecules. The destruction of ozone occurs mainly through catalytic reactions with other gases, such as chlorine and bromine. The total amount of ozone in the stratosphere will remain fairly constant (relative to the well-recognized seasonal variations) as long as there is no change in the destruction rate and the transport of ozone out of the stratosphere. However, increasing inputs of chlorine and bromine into the stratosphere over the past few decades have changed this balance.

Measurements of ozone by satellite and ground-based instruments over the last several decades indicate that stratospheric ozone levels have decreased significantly. Amounts of ozone in the global atmosphere have decreased globally by more than 5% since 1970. Figure 5 shows satellite and ground-based measurements of the change in ozone from 60°S to 60°N latitudes. Satellite measurements began in late 1978, while a reasonably representative global network of ground-based sta-

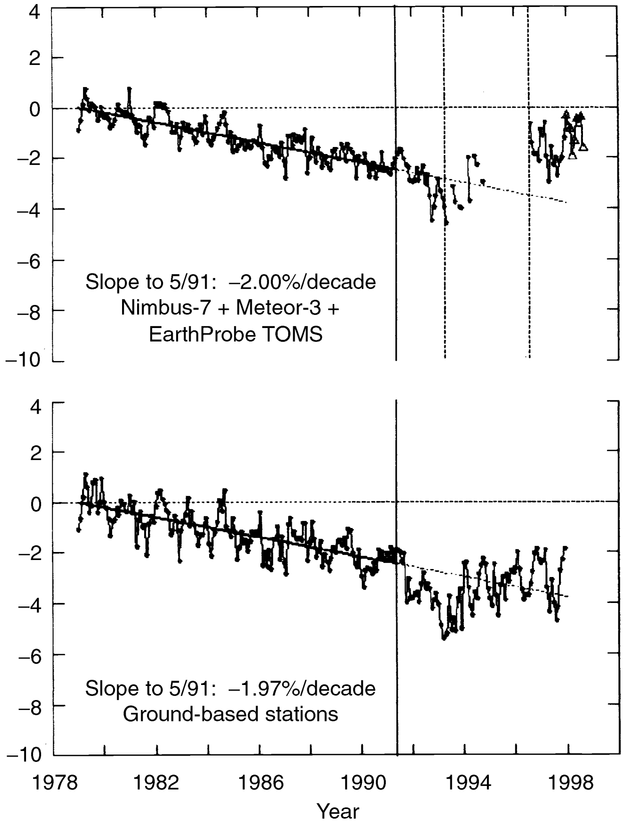


FIGURE 5 Deviations in total ozone with time relative to January 1979 from various ground-based (bottom panel) and satellite measurements (top panel). The data are area-weighted over 60° South to 60° North. Based on the World Meteorological Organization 1998 assessment of ozone depletion.

tions has been in operation since the late 1960s. In the data shown here, the seasonal variations and the effects of the variations in the solar flux due to the 11-year solar cycle have been removed (which would produce about a 1.2–1.5% variation from solar minimum to solar maximum).

Significant decreases in the total (integrated column of) ozone are found in both the Northern and Southern Hemispheres at the middle and high latitudes, with no significant change in the tropics. Much larger decreases in total ozone are found at latitudes greater than 60°, particularly in the Southern Hemisphere, as discussed later. Measurements indicate that the total ozone column at midlatitudes in the Northern Hemisphere has decreased by 1.3% per decade in the summer months and 2.7% per decade in the winter months since 1969. Satellite and ozonesonde (balloon) data sets indicate that ozone is particularly decreasing in the lower stratosphere, accounting for a major fraction of the trend in total ozone, although there is also significant ozone destruction occurring in the upper stratosphere. At the end of the ozone record in Fig. 5, there is an upturn in ozone associated with the recovery from the effects of the Mount Pinatubo volcanic eruption in 1991. Also, in the top part of Fig. 5, the switch to the new satellite in 1996 resulted in a bias that was not yet corrected at the time of the international assessment in 1998. When this article was written, there was no indication of a recovery in ozone due to the control measures on CFCs and other halocarbons.

Beginning in the late 1970s, a special phenomenon began to occur in the springtime over Antarctica, referred to as the Antarctic ozone "hole." A large decrease in the total ozone, now over a 60% decrease relative to pre-hole levels, has been observed in the springtime (September to November) above Antarctica. Joe Farman of the British Antarctic Survey and co-authors documented this rapid springtime decrease in Antarctic ozone over the ozone measurement station at Halley Bay, Antarctica, attracting the attention of the scientific community. Decreases in the total ozone column of more than 50% were found compared with historical values observed by both ground-based and satellite techniques. Measurements made in 1987 indicated that more than 95% of the ozone over Antarctica at altitudes from 13 to 22 km had disappeared during September and October. More recently, the Antarctic ozone holes since 1992 have been quite comparable, being the biggest (areal extent) and the deepest (minimum amounts of ozone overhead), with ozone being locally depleted by more than 99% between about 14–19 km in October.

B. Causes and Consequences of Stratospheric Ozone Depletion

The connection between potential environmental effects and man-made chlorofluorocarbons was first raised by Drs. Mario Molina and F. Sherwood Rowland in 1974, when they suggested that chlorine from these compounds could destroy stratospheric ozone. Research findings since then have continued to support the significant role that these compounds play in affecting the global distribution of ozone. In addition, it is recognized that other gases containing chlorine and bromine, which are even more reactive with O_3 than chlorine, are also affecting ozone. To a lesser extent, the increasing concentrations of other gases like CO_2, CH_4, and N_2O are also affecting stratospheric ozone.

The inverse relationship between changes in ozone and UV-B is well established by both theoretical analyses and observations. A number of studies have shown that the corresponding increase in UV-B at the ground resulting from ozone depletion can lead to increased incidences of skin cancers, cataracts, and other effects on humans and animals.

The recognition of the deleterious effect of chlorine and bromine on ozone spawned international action to restrict the production and use of CFCs and halons to protect stratospheric ozone. These included the 1987 Montreal Protocol on Substances That Deplete the Ozone Layer and the subsequent 1990 London Amendment, 1992 Copenhagen Amendment, and the 1997 Montreal Amendment. The agreements initially called for reduction of CFC consumption in developed countries. A November 1992 meeting of the United Nations Environment Programme held in Copenhagen resulted in substantial modifications to the protocol because of the large observed decrease in ozone, and called for the phase-out of CFCs, carbon tetrachloride (CCl_4), and methyl chloroform (CH_3CCl_3) by 1996 in developed countries. As part of this, the United States, through the Clean Air Act, has also eliminated the production and import of these chemicals. Production of these compounds is to be totally phased out in developing countries by 2010, while production of halons in developed countries was stopped in 1994. Human-related production and emissions of methyl bromide are not to increase after 1994 in developed countries, with total elimination by 2005.

C. Projected Trends in Ozone

Projected changes in globally averaged total ozone, using several assumptions about future emissions of CFCs, halons, and their replacements, have been evaluated using models of atmospheric processes. With the original Montreal Protocol provisions, there would still have been a significant reduction in total ozone, as much as 15% by 2050 relative to 1980 levels, according to some analyses. It is only under the London and Copenhagen Amendments provisions, which call for the complete phase-out of CFCs and halons and other halocarbons, that the ozone reduction trend is reversed. The largest ozone reductions are reached at about 1998 to 2000. After this, ozone begins to recover, although it is not until about 2050 that the 1980 level of global total ozone would again be expected. Thus, according to these models, it will roughly be the middle of the next century before the chlorine and bromine in the stratosphere are reduced to levels corresponding to those when the Antarctic ozone "hole" first began. Also, without any control measures and assuming unfettered growth, the global mean total column ozone would decrease about 30% by 2050, decreasing further with time.

IV. AIR POLLUTANTS IN THE LOWER ATMOSPHERE

The elevated concentration of oxidants in urban regions is a difficult problem in many parts of the world. In the United States, progressively tighter emissions controls have been implemented since the late 1970s to control production of ozone. Most of these controls have been aimed at emissions of NO_x and VOCs, particularly the former. Nonetheless, many urban areas still fail to meet the federal standards. In addition, there is growing concern that enhanced levels of ozone concentrations in rural areas downwind of urban areas are resulting in harmful effects on agricultural and forest ecosystems. Part of the difficulty is understanding how emissions of hydrocarbons from the biosphere affect ozone levels in urban areas.

V. THE FUTURE AND ATMOSPHERIC GASES

Changes in climate associated with increasing concentrations of greenhouse gases could have significant effects on the biosphere and on biodiversity. Corresponding changes in the biosphere could affect biogeochemical cycles and the sources and sinks of atmospheric gases, leading to further changes in the climate. Such interactions with the biosphere and with biodiversity are inadequately considered in current climate projections. The effects of such interactions require much more study.

Stratospheric ozone levels should largely recover over the next half century. However, the timing of this recovery will largely depend on the emissions of halocarbons and on the emissions of other gases that influence stratospheric chemistry. Changes in the oxidative capacity could also change the amounts of gases reaching the stratosphere and thus the rate of and extent of recovery. Climate change would also likely influence stratospheric temperature and winds, thus further affecting the rate of ozone recovery. The long-term impacts of the reduced levels of ozone on the biosphere and on biodiversity remain poorly understood.

Attempts to control pollution in urban regions continue to increase. However, there is still much to learn before such controls can be fully effective. The U.S. National Academy of Sciences has recently stated that too little is known about the transport and deposition of atmospheric gases, including toxics and nutrients, to the biosphere and their interaction with biota. Further study of the rates of chemical exchange between the atmosphere and ecosystems will provide better quantitative estimates of atmospheric chemical impacts on the biosphere and biospheric emissions to the atmosphere.

See Also the Following Articles

AIR POLLUTION • CARBON CYCLE • CLIMATE CHANGE AND ECOLOGY, SYNERGISM OF • CLIMATE, EFFECTS OF • GREENHOUSE GASES • NITROGEN AND NITROGEN CYCLE • ULTRAVIOLET RADIATION

Bibliography

Brasseur, G. P., Orlando, J. J., and Tyndall, G. S. (eds.). (1999). *Atmospheric Chemistry and Global Change.* Oxford University Press, New York.

Graedel, T. E., and Crutzen, P. J. (1993). *Atmospheric Change: An Earth System Perspective.* W. H. Freeman and Company, New York.

Hewitt, C. N., and Sturges, W. T. (ed.). (1993). *Global Atmospheric Chemical Change.* Elsevier Applied Science, New York.

Intergovernmental Panel on Climate Change. (1996). *Climate Change 1995: The Science of Climate Change.* Cambridge University Press, New York.

National Research Council. (1998). *The Atmospheric Sciences Entering the Twenty-First Century.* National Academy Press, Washington, D.C.

Singh, H. B. (ed.). (1995). *Composition, Chemistry, and Climate of the Atmosphere.* Van Nostrand-Reinhold, New York.

Turco, R. P. (1997). *Earth under Siege: From Air Pollution to Global Change.* Oxford University Press, New York.

World Meteorological Organization. (1995). *Scientific Assessment of Ozone Depletion: 1994.* Global Ozone Research and Monitoring Project, Report No. 37. World Meteorological Organization, Geneva.

Wuebbles, D. J., Jain, A., Edmonds, J., Harvey, D., and Hayhoe, K. (1999). Global change; The state of the science. *Environ. Pollution* 100, 57–86.

AUSTRALIA, ECOSYSTEMS OF

Raymond L. Specht* and Alison Specht†
*The University of Queensland and †School of Resource Science and Management, Southern Cross University

GLOSSARY

biogeographic region A region that contains a “homogeneous” suite of Floristic Groups sorted by some classificatory program.

community diversity Number of ecosystems and associated plant and animal species per 1° latitude by 1° longitude grid-cell.

diagnostic species Species that emerge from semiquantitative analysis as being the key species associated with a floristic group.

evaporative aerodynamics Study of the influence of aerodynamics on the development of both leaf structure and foliage distribution throughout a plant community.

floristic group A plant community that contains a “homogeneous” suite of species sorted by some classificatory program.

foliage projective cover (FPC) The horizontal cover of leaves (in crowns of varying foliage density) in overstory and understory strata of an open-structured plant community, measured using vertical cross-wire sighting tubes.

shoot growth Biomass production (per hectare) of foliage shoots produced annually in the overstory of a plant community.

species richness Number of species of plant or animal per hectare.

DISTINCTIVE ECOSYSTEMS FOUND IN AUSTRALIA range from the arid lands to the perhumid evaporative zone in both the temperate south and the tropical north of the continent. After applying the classificatory program TWINSPAN to almost 5000 ecological surveys, the floristic components of Australia’s plant communities were sorted into 338 Floristic Groups. Eight biogeographic regions, with 45 subdivisions, were defined. The species richness (the number of species per hectare) of overstory plants is related to the annual biomass production of that stratum. Because annual growth of the understory stratum depends on the transmission of solar radiation through the overstory, species richness of this stratum is higher in tropical than in temperate climates, and in the early stages of secondary succession when the overstory stratum is poorly developed. The species richness of non-arboreal vertebrates parallels that of the understory stratum; species richness of epi-

gaeic invertebrates depends on the annual supply of leaf litter, largely from the overstory. The diversity of Floristic Groups (per 1° latitude by 1° longitude grid-cell) is related to the annual biomass production of overstory foliage in each evaporative climate. As annual shoot growth is markedly reduced by the low-phosphate status of the relictual Gondwanan soils of the northern, central, and western parts of the continent, the diversity of major plant communities in these areas is lower than that found on the richer soils of the east.

I. AUSTRALIAN ECOSYSTEMS

Australia is a continental mass lying south of the equator between latitudes 11° and 44°. The climate varies markedly with geographical position: summer rainfall predominates in the north, winter rainfall is characteristic of the south, and the eastern seaboard shows a more general distribution of rainfall throughout the year. Marked evaporative gradients from the humid coastal fringe to the arid inland are found in the south, east, and north of the continent. The structure of the vegetation varies along these evaporative gradients: from the dense rainforests in eastern Australia, to the tall eucalypt forests of the south-east and south-west, to the eucalypt forests/woodlands, with either grassy or heathy understory, distributed throughout the continent. Heathlands and grasslands are found in tropical to temperate climes, the mallee eucalypt vegetation is found on the calcareous soils of the south, and a variety of other vegetation structures are found in the vast arid zone and in the alpine, coastal, and wetland landscapes of the continent.

Most families and many genera now found in the Australian flora developed over 100 million years ago, before the Gondwanan supercontinent began to break up. Vast expanses of infertile lateritic soils developed in the humid/perhumid climate of that time. The onset of aridity in the early Tertiary, some 50 million years ago, caused fragmentation of the humid/perhumid Gondwanan vegetation, enabling its less humid elements to expand and to evolve, eventually to cover much of the continent.

The vegetation of today comprises those taxa that were able to survive as the evaporative climate became drier. A distinctive heathy vegetation developed on the nutrient-poor soils of lateritic origin, characteristic of a large part of the original Gondwanan super-continent. The attributes that enabled most of the Gondwanan flora to survive over the last 50 million years also facilitates their rapid regeneration from underground organs and epicormic buds whenever a disturbance such as a fire occurs.

II. EVAPORATIVE AERODYNAMICS

As most of the continent of Australia experiences a dry season, both the distribution of foliage in the overstory and understory and its eco-physiological attributes result from the growth of a multitude of shoot apices (in both overstory and understory strata) in equilibrium with the evaporative aerodynamics of the atmosphere flowing over and through the plant community (Specht, 1972; Specht and Specht, 1999). The amount of available water affects the vertical growth of each foliage shoot (and the number of essentially identical leaves produced annually), but not the horizontal cover of the foliage throughout the plant community (Specht and Specht, 1989a); this is determined by the evaporative power of the atmosphere, a relatively constant factor from year to year.

After any perturbation, the horizontal cover of the foliage and its eco-physiological characteristics are rapidly restored in both the overstory and understory of an evergreen plant community. The ratio of the actual to potential evapotranspiration (per hectare) from the plant community is linearly related to the amount of water available during each month; the slope of this linear regression is called the Evaporative Coefficient (Specht, 1972). Values of this community-physiological constant have been computed for meteorological stations throughout Australia and the resulting isolines are mapped in Fig. 1.

III. MAJOR PLANT COMMUNITIES IN AUSTRALIA

The great number of tree species (20 to 140) that coexist in subtropical and tropical plant communities in Australia, from the arid to the perhumid evaporative zones, makes the definition of statistically "homogeneous" suites within these vegetations a challenge, one that is not faced by temperate plant ecologists who deal with simple systems with relatively low numbers of overstory species per hectare (1 to 10 in number). This has necessitated the adoption of computer techniques.

TWINSPAN Floristic Groups have been defined in all Australian plant formations (Specht *et al.*, 1995). To accomplish this, species lists reported for plant communities throughout the continent over the last 70 years

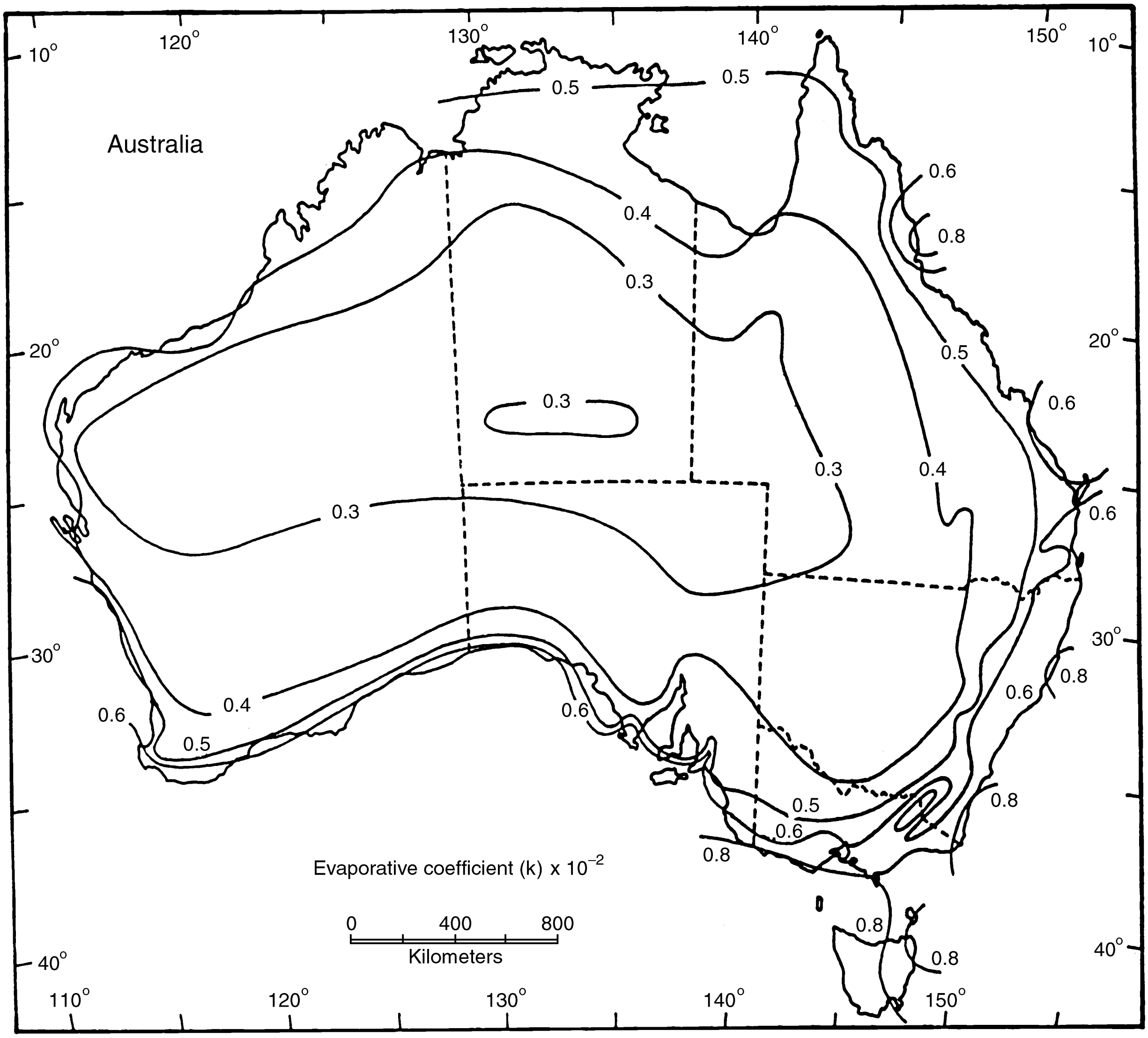


FIGURE 1 Map of Australia showing isolines of the Evaporative Coefficient (k). Climatic zones may be defined as follows:—Arid Zone, $k < 0.35$; Semi-arid Zone, $k \equiv 0.35–0.45$; Subhumid Zone, $k \equiv 0.45–0.55$; Humid Zone, $k \equiv 0.55–0.75$; Perhumid Zone, $k > 0.75$. (All values of k are $\times 10^{-2}$ per mm of soil water available per month.) (After Specht, 1972; Specht and Specht, 1999)

were collated in 16 large data banks and arranged according to structural formation (Table I). These formations were then analyzed by TWINSPAN (M. O. Hill, *Ecology* **61**, 237–249, 1973) on the basis of the presence or absence of a species in each list.

TWINSPAN first sorted tropical/subtropical plant communities in northern Australia from the temperate/montane plant communities in southern Australia. The program then sorted the tropical from the subtropical, and the temperate from the montane. Further subdivision separated the humid plant communities from the subhumid, and so on for plant communities that had developed in different micro-environments. The number of TWINSPAN Floristic Groups and their subdivisions that were recorded in each 30′ latitude by 30′ longitude grid-cell are shown in the *Conservation Atlas of Plant Communities in Australia* (Specht *et al.*, 1995). The diversity of Floristic Groups is greatest in the perhumid zone, and least in the arid evaporative zone.

TWINSPAN analysis of the sixteen data banks resulted in the definition of 338 overstory Floristic Groups. A further 60 understory Floristic Groups have been recognized in open-structured plant communities (Specht *et al.*, 1995).

TABLE I

Structural Characteristics of Major Plant Formations in Australia[a]

Vegetation formation	Climatic zone	Overstory: Stand height (m)	Overstory: Foliage projective cover (%)	Overstory: Leaf specific weight[b] (mg cm^{-2})	Understory
Overstory with trees					
Closed-forest	Perhumid	20[c]–10	100–70	8–16	Regeneration saplings, shade plants
Open-forest	Humid	>30–10	70–30	10–28	Grassy, heathy
Woodland	Subhumid	20–10	30–10	13–40	Grassy, heathy
Open-scrub	Semiarid	5–2	70–30	18–50	Grassy, heathy
Tall shrubland	Arid	5–2	30–10	20–>50	Grassy, hummock grass, chenopods
Treeless vegetation					
Heathland	Humid–subhumid	2–0.25	70–30	Sclerophyllous	—
Dwarf heathland (fellfield)	Montane	<0.25	30–<10	Sclerophyllous	—
Tussock grassland	Humid–subhumid	1–0.25	70–30	Herbaceous	—
Open tussock grassland	Semiarid	<0.25	30–10	Herbaceous	—
Hummock grassland	Semiarid–arid	1–0.5	70–10	Sclerophyllous	—
Chenopod low shrubland	Arid	2–0.25	30–10	Semisucculent	—

[a] From Specht *et al.* (1995) and Specht and Specht (1999).

[b] Leaf/leaflet area decreases, and leaf specific weight increases, in the 10°C temperature gradient from tropical to subtropical to temperate regions.

[c] In the tropics and subtropics, these are emergents and strangler figs, or lianas and epiphytes.

IV. DIAGNOSTIC SPECIES IN AUSTRALIAN FLORISTIC GROUPS

The plant species which define the floristic groups (Specht *et al.*, 1995) determined by TWINSPAN in the 29 Australian plant formations (Table I) are listed below. Each floristic group has been given a code: capital letters for the formation (e.g., RF = closed-forest (rainforest), T = open-forest and woodland), followed by numerals. The numerals refer to the dichotomous subdivisions in the TWINSPAN classification; within a formation, groups having consecutive odd–even pairs of numerals are most closely related (e.g., RF35 and RF36). In species-rich formations, such as closed-forests, subdivisions of the primary TWINSPAN groups are shown by the numbers in parentheses.

Closed-forests (rainforests)—tropical north-eastern Australia (after L. J. Webb, J. G. Tracey, and W. T. Williams, *Aust. J. Ecol.* **9**, 169–198, 1984)

RF35(140) *Drypetes deplanchei*
RF35(141) *Endiandra glauca - Polyscias australiana*
RF35(142) *Calophyllum australianum - Castanospermum australe*
RF35(143) *Polyalthia nitidissima*
RF36(144) *Agathis robusta - Diospyros pentamera*
RF36(145) *Myristica insipida*
RF36(146) *Polyscias australiana - Podocarpus neriifolius*
RF36(147) *Archontophoenix alexandrae - Ficus racemosa*
RF37(75) *Acronychia acidula - Dendrocnide photinophylla*
RF37(148) *Archontophoenix alexandrae - Castanospermum australe*
RF37(149) *Musgravea heterophylla - Normanbya normanbyi*
RF38(76) *Beilschmiedia tooram - Endiandra sankeyana*
RF38(155) *Cryptocarya angulata - C. corrugata*
RF38(308) *Cryptocarya mackinnoniana - Flindersia bourjotiana*
RF38(309) *Balanops australiana - Lomatia fraxinifolia*
RF39 *Laccospadix australasica - Steganthera macooraia*

Closed-forests (rainforests)—subtropical eastern Australia (after L. J. Webb, J. G. Tracey, and W. T. Williams, *Aust. J. Ecol.* **9**, 169–198, 1984)

RF40(160) *Austromyrtus acmenoides - Litsea leefeana-Sloanea woollsii*
RF40(163) *Helicia ferruginea - Litsea leefeana*

RF40(322) *Rhodamnia argentea - Syzygium francisii*
RF40(323) *Cryptocarya obovata - Sloanea woollsii*
RF40(324) *Acmena ingens - Zanthoxylum brachyacanthum*
RF40(325) *Schizomeria ovata - Argyrodendron actinophyllum*
RF41(165) *Cupaniopsis anacardioides*
RF41(167) *Celtis paniculata - Cryptocarya triplinervis*
RF41(328) *Dendrocnide* spp. - *Ficus macrophylla - Toona ciliata*
RF41(329) *Acmena smithii - Ficus coronata - Toona ciliata*
RF41(332) *Cryptocarya glaucescens - Mischocarpus pyriformis*
RF42(168) *Orites excelsa - Sloanea woollsii*
RF42(169) *Nothofagus moorei - Cryptocarya foveolata - Orites excelsa*
RF42(170) *Nothofagus moorei - Tristaniopsis laurina*
RF42(171) *Doryphora sassafras*
RF43(86) *Acmena smithii - Ficus coronata - Eupomatia laurina*
RF43(87) *Acacia melanoxylon - Acmena smithii*
RF44(88) *Cryptocarya floydii - Pittosporum undulatum*
RF44(178) *Podocarpus elatus - Syzygium oleosum*
RF44(359) *Araucaria cunninghamii - Flindersia australis*
RF44(358) *Araucaria cunninghamii - Argyrodendron trifoliolatum - Drypetes* sp.
RF45(180) *Araucaria cunninghamii - Argyrodendron trifoliolatum - Cryptocarya bidwillii*
RF45(181) *Murraya ovatifoliolata*
RF45(182) *Araucaria cunninghamii - Austromyrtus bidwillii*
RF45(183) *Strychnos axillaris*
RF46(93) *Apophyllum anomalum - Maytenus cunninghamii*
RF46(184) *Cupaniopsis anacardioides - Pleiogynium timorense*
RF46(185) *Casearia multinervosa - Croton insularis*

Closed-forests (rainforests)—temperate south-eastern Australia (after L. J. Webb, J. G. Tracey, and W. T. Williams, *Aust. J. Ecol.* **9**, 169–198, 1984)

RF7 *Nothofagus cunninghamii - Athrotaxis cupressoides - Diselma archeri*
RF13 *Nothofagus cunninghamii - Atherosperma moschatum*
RF24 *Nothofagus cunninghamii - Pittosporum bicolor*
RF25 *Atherosperma moschatum - Pittosporum bicolor*

Semi-deciduous closed-forests—monsoonal northern Australia (J. Russell-Smith, *J. Veg. Sci.* **2**, 259–278, 1991)

MRF1 *Aglaia sapidina - Hydriastele wendlandiana*
MRF2 *Acacia auriculiformis - Carpentaria acuminata*
MRF3 *Acmenosperma claviflorum - Macaranga involucrata*
MRF4 *Calophyllum sil - Ficus hispida - Ilex arnhemensis*
RF34(69) *Carallia brachiata - Ilex arnhemensis*
MRF5 *Acmenosperma claviflorum - Litsea breviumbellata*
MRF6 *Omalanthus novo-guineensis - Xanthostemon eucalyptoides*
MRF7 *Allosyncarpia ternata - Calophyllum sil*
MRF8 *Allosyncarpia ternata - Alyxia ruscifolia*
MRF9 *Aglaia rufa - Diospyros maritima - Secamone elliptica*
MRF10 *Barringtonia acutangula - Livistona benthamii*
MRF11 *Ficus racemosa - Melaleuca argentea*
MRF12a *Canarium australianum - Celtis philippensis*
MRF12b *Alstonia spectabilis - Denhamia obscura*
MRF13 *Ficus coronulata - F. virens - Glochidion apodogynum*
MRF14 *Brachychiton collinus - Callicarpa candicans*
MRF15 *Brachychiton spectabilis - Ziziphus quadrilocularis*
RF47 *Brachychiton chillagoensis - Terminalia aridicola*
RF16(132) *Gyrocarpus americanus - Lysiphyllum cunninghamii*
RF16(32) *Ficus platypoda* var. *minor*

Semi-deciduous closed-forests—subtropical eastern Australia (P. I. Forster, P. D. Bostock, *et al.*, Queensland Herbarium, Brisbane, 1991)

DS8(16) *Cupaniopsis parvifolia - Diospyros humilis - Owenia venosa*
DS8(17) *Cupaniopsis parvifolia - Maclura cochinchinensis*
DS9(18) *Archidendropsis thozetiana - Cupaniopsis wadsworthii*
DS9(19) *Archidendropsis thozetiana - Brachychiton rupestris*
DS10(20) *Pleiogynium timorense - Pouteria sericea*
DS10(21) *Alchornia ilicifolia - Cryptocarya triplinervis*
DS11(22) *Cupaniopsis parvifolia*
DS11(23) *Cupaniopsis parvifolia - Araucaria cunninghamii*
DS12(24) *Mischocarpus pyriformis - Alyxia ruscifolia*
DS12(25) *Mischocarpus pyriformis - Araucaria cunninghamii*
DS13(26) *Araucaria cunninghamii - Alectryon tomentosus - Cleistanthus cunninghamii*
DS13(27) *Araucaria cunninghamii - Alectryon tomentosus - A. subcinereus*
DS14(28) *Cryptocarya macdonaldii - C. laevigata*
DS14(29) *Cryptocarya macdonaldii*
DS15(30) *Araucaria cunninghamii - Calamus muelleri - Wilkiea macrophylla*

DS15(31) *Araucaria cunninghamii - Maclura cochinchinensis - Smilax australis*
DS40 *Alectryon forsythii - A. subdentatus - Notelaea microcarpa*

Eucalypt open-forests and woodlands—monsoonal northern Australia (A. N. Gillison *Ecosystems of the World*, Vol. 13, pp. 183–243. Elsevier, Amsterdam, 1983; R. L. Specht and D. W. Drake, unpubl. data)

T187a *Eucalyptus leucophloia*
T189 *Eucalyptus microneura*
T359 *Eucalyptus brownii - E. crebra*
T368 *Eucalyptus cullenii - E. shirleyi - Corymbia erythrophloia*[1]
T369 *Corymbia papuana - C. erythrophloia*
T371 *Corymbia papuana - C. polycarpa - Eucalyptus platyphylla*
T372 *Eucalyptus tectifica - Corymbia grandifolia - C. foelscheana*
T373a *Eucalyptus tetrodonta - E. miniata - Corymbia ferruginea*
T738 *Eucalyptus whitei - Corymbia papuana*
T739 *Coymbia papuana - C. erythrophloia*

Eucalypt open-forests and woodlands—subtropical eastern Australia (R. L. Specht and D. W. Drake, unpubl. data)

T168 *Eucalyptus andrewsii - E. youmanii - E. cameronii*
T168/343 *Eucalyptus deanei - E. campanulata*
T346 *Euc. campanulata - E. laevopinea - E. cypellocarpa - E. obliqua*
T346b *Eucalyptus saligna - E. microcorys*
T351 *Eucalyptus campanulata - E. saligna*
T352 *Eucalyptus amplifolia - Angophora subvelutina*
T360 *Eucalyptus chloroclada - Angophora costata* subsp. *leiocarpa*
T361 *Eucalyptus chloroclada - E. sideroxylon*
T362 *Eucalyptus microcarpa*
T363a *Eucalyptus microcarpa - E. populnea*
T364 *Eucalyptus populnea - E. orgadophila*
T365 *Eucalyptus melanophloia*
Calcolu *Callitris glaucophylla*
Serp-R *Eucalyptus fibrosa* subsp. nov. *Corymbia xanthope - Macrozamia* sp. nov.
Serp-W *Eucalyptus crebra - Corymbia erythrophloia - Macrozamia* sp. nov.
Serp-B *Eucalyptus ophitica - Xanthorrhoea glauca*
T706 *Eucalyptus pilularis - Corymbia gummifera - Angophora* spp.
T707 *Eucalyptus paniculata - E. propinqua*
T710 *Euc. acmenoides - E. pilularis - E. microcorys - Syncarpia glomulifera*
T710b *Eucalyptus grandis*
Eucbiturb *Eucalyptus biturbinata*
Euccloe *Eucalyptus cloeziana - E. tenuipes*
Shill *Syncarpia hillii*
T711 *Eucalyptus signata - Corymbia intermedia*
T712 *Eucalyptus crebra - Corymbia citriodora - C. trachyphloia*
T713 *Corymbia intermedia - Lophostemon* spp.
T714 *Eucalyptus tereticornis - Lophostemon suaveolens*
T715 *Eucalyptus tereticornis - Corymbia tessellaris - C. clarksoniana*
T716a *Eucalyptus crebra - E. tereticornis - Corymbia maculata*
T716b *Eucalyptus moluccana - E. tereticornis*
T717 *Eucalyptus crebra - Corymbia maculata*

Eucalypt open-forests and woodlands—temperate south-eastern Australia (R. L. Specht and D. W. Drake, unpub1. data)

T172 *Eucalyptus regnans*
T328a *Eucalyptus odorata - E. porosa*
T329 *Eucalyptus leucoxylon*
T330 *Eucalyptus fasciculosa - E. leucoxylon*
T331 *Eucalyptus baxteri - E. fasciculosa*
Eucpryor *Eucalyptus pryoriana - E. viminalis* subsp. *cygnetensis*
Eucnit *Eucalyptus nitida*
T338 *Eucalyptus albens - E. blakelyi - E. melliodora*
T339 *Eucalyptus dealbata - E. blakelyi - E. sideroxylon*
T340 *Eucalyptus blaxlandii - E. goniocalyx*
T341 *Eucalyptus macrorhyncha - E. polyanthemos*
T342 *Eucalyptus dives - E. macrorhyncha - E. radiata*
T347 *Eucalyptus cypellocarpa - E. obliqua - E. radiata*
T348 *Eucalyptus globoidea - E. sieberi*
T349 *Eucalyptus globoidea - Corymbia gummifera*
T350 *Eucalyptus piperita - E. sieberi-Corymbia gummifera*
T40 *Eucalyptus cordata*
T664 *Eucalyptus nitida*
T665 *Eucalyptus cephalocarpa - E. globoidea*
T666 *Eucalyptus viminalis*
T667 *Eucalyptus obliqua*

Eucalypt open-forests and woodlands—montane south-eastern Australia (R. L. Specht and D. W. Drake, unpubl. data)

T8 *Eucalyptus coccifera*
T36 *Eucalyptus obliqua - E. amygdalina - E. globulus*
T38 *Eucalyptus delegatensis - E. dalrympleana - E. viminalis*
T75 *Eucalyptus urnigera - E. delegatensis*

---

[1] *Corymbia* was formerly regarded as a subgenus of *Eucalyptus*.

T79 *Eucalyptus brookeriana - E. delegatensis*
T148 *Eucalyptus obliqua*
T149 *Eucalyptus delegatensis - E. regnans - E. obliqua*
T156 *Eucalyptus delegatensis - E. pauciflora - E. rubida*
T157 *Eucalyptus delegatensis - E. pauciflora - E. rodwayi*
T334 *Eucalyptus pauciflora - E. rubida*
T335 *Eucalyptus pauciflora - E. stellulata*

Eucalypt open-forests and woodlands—temperate south-western Australia (R. L. Specht and A. J. M. Hopkins, unpubl. data)

SW20 *Eucalyptus gomphocephala - Agonis flexuosa*
SW26 *Eucalyptus loxophleba*
SW43a *Eucalyptus marginata - Corymbia calophylla*
SW47 *Eucalyptus marginata - Banksia* spp. *- Allocasuarina humilis*
SW85 *Eucalyptus diversicolor*
SW96a *Eucalyptus wandoo*
SW97a *Eucalyptus salmonophloia*
SW100 *Eucalyptus loxophleba - E. salmonophloia*

Eucalypt open-forests and woodlands—Australian wetland forests (R. L. Specht, *Ecosystems of the World*, Vol. 16, pp. 387–406. Elsevier, Amsterdam, 1990; R. L. Specht and D. W. Drake, unpubl. data)

T188 *Eucalyptus coolabah, E. microtheca,* etc.
T191 *Eucalyptus ochrophloia - E. coolabah*
T190 *Eucalyptus camaldulensis*
T352b *Casuarina cunninghamiana*
T354 *Melaleuca quinquenervia - Eucalyptus robusta*
T366a *Eucalyptus largiflorens*
T366b *Allocasuarina luehmannii*
T370 *Eucalyptus leptophleba*
SW4a *Eucalyptus rudis - Melaleuca rhaphiophylla*
SW4b *Eucalyptus rudis - Corymbia calophylla*
SW84 *Agonis hypericifolia - Banksia littoralis*
T801 *Melaleuca leucadendra - M. viridiflora*

Mallee eucalypt open-scrubs—monsoonal northern Australia (R. L. Specht, unpubl. data)

Euccupu *Eucalyptus cupularis*
Eucmann *Eucalyptus mannensis*
Eucnorm *Eucalyptus normantonensis*
Eucpers *Eucalyptus persistens*
Eucoxym *Eucalyptus oxymitra*
Eucpach *Eucalyptus pachyphylla*
Eucset *Corymbia setosa*
Eucthoz *Eucalyptus thozetiana - Acacia catenulata - A. aneura*

Mallee eucalypt open-scrubs—subtropical eastern Australia (R. L. Specht, unpubl. data)

Eucappr *Eucalyptus approximans*
Eucbake *Eucalyptus bakeri*
Euccurt *Eucalyptus curtisii*
Serp-GW *Eucalyptus serpenticola - E. nortonii - Allocasuarina ophiolitica*

Mallee eucalypt open-scrubs—temperate south-eastern Australia (R. L. Specht, *Ecosystems of the World*, Vol. 11, pp. 203–231. Elsevier, Amsterdam, 1981)

M7 *Eucalyptus diversifolia - E. cosmophylla - E. rugosa - Melaleuca uncinata*
M10 *Eucalyptus behriana*
M13 *Eucalyptus diversifolia - E. socialis*
M23 *Eucalyptus socialis - E. dumosa - E. gracilis - E. polybractea*
M24 *Eucalyptus incrassata - E. viridis - E. polybractea*
M25 *Eucalyptus incrassata - E. foecunda - Melaleuca uncinata*
M4 *Eucalyptus socialis - E. dumosa*
M44 *Eucalyptus socialis - E. dumosa - E. gracilis - Acacia rigens*
M45 *Eucalyptus socialis - E. dumosa - E. gracilis - Enchylaena tomentosa*
Eucinte *Eucalyptus intertexta - Callitris glaucophylla*
Euckits *Eucalyptus kitsoniana*
Eucobtus *Eucalyptus obstans - E. luehmanniana*

Mallee eucalypt open-scrubs—temperate south-western Australia (R. L. Specht and A. J. M. Hopkins, unpubl. data)

SW27 *Eucalyptus diversifolia - E. oleosa - Melaleuca lanceolata*
SW51 *Eucalyptus leptopoda - Callitris preissii - Melaleuca cordata*
SW101b *Eucalyptus eremophila - E. forrestiana - E. oleosa*
SW196 *Allocasuarina campestris*
SW197 *Allocasuarina campestris - Acacia acuminata - Eucalyptus obtusiflora*

Heathlands and related shrublands—monsoonal northern Australia (R. L. Specht and W. E. Drake, unpubl. data)

H7 *Corymbia ferruginea - Jacksonia odontoclada - Plectrachne pungens*
H7a *Calytrix exstipulata - Grevillea pteridifolia - Persoonia falcata*
H12 *Asteromyrtus lysicephala - Jacksonia thesioides*
H13 *Neofabricea myrtifolia - Neoroepera banksii*
Baecfrut *Baeckea frutescens*
Leptpurp *Leptospermum purpurascens*

Heathlands and related shrublands—subtropical eastern Australia (R. L. Specht, *Ecosystems of the World*, Vol. 9A, pp. 125–210. Elsevier, Amsterdam, 1979; R. L. Specht and W. E. Drake, unpubl.)

H40 *Allocasuarina littoralis - Alphitonia excelsa*

H82 *Melaleuca nodosa - Gahnia sieberana - Blechnum indicum*
H166 *Leptospermum polygalifolium - Phyllota phylicoides*
H167 *Leptospermum polygalifolium - Platysace linearifolia*
Xanth *Xanthorrhoea* spp.

Heathlands and related shrublands—temperate south-eastern Australia (R. L. Specht and W. E. Drake, unpubl. data)

H44a *Melaleuca uncinata*
H44b *Banksia ornata - Xanthorrhoea caespitosa - Allocasuarina pusilla*
H45 *Melaleuca uncinata - Xanthorrhoea semiplana* var. *tateana*
H84 *Banksia ericifolia - Hakea dactyloides*
H170 *Banksia marginata - Epacris impressa*
H171 *Banksia marginata - Epacris impressa - Pteridium esculentum*

Heathlands and related shrublands—montane south-eastern Australia (J. B. Kirkpatrick, *The Scientific Significance of the Australian Alps,* pp. 127–142. Aust. Acad. Sci., Canberra, 1989)

AV1 *Carex gaudichaudiana - Myriophyllum pedunculatum*
AV2 *Carex gaudichaudiana*
AV3 *Baeckea gunniana - Epacris breviflora - Carex gaudichaudiana*
AV4 *Baeckea gunniana - Epacris paludosa - E. lanuginosa*
AV5 *Oxylobium alpestre - Olearia phlogopappa - Phebalium ovatifolium*
AV6 *Grevillea australis - Epacris paludosa - Poa* spp.
AV7 *Leucopogon montanus - Poa* spp.
AV8 *Podocarpus lawrencei*
AV9 *Microtrobos niphophilus - Epacris serpyllifolia*
AV10 *Abrotanella forsterioides - Richea scoparia*
AV11 *Carpha rodwayi - Oreobolus pumilio - Phyllachne colensoi*
AV12 *Diselma archeri - Microcachrys tetragona*
AV13 *Gaultheria depressa - Senecio leptocarpus*
AV14 *Epacris serpyllifolia - Orites revoluta - Richea sprengelioides*
AV15 *Epacris serpyllifolia - Richea scorparia - Senecio leptocarpus*
AV16 *Eucalyptus vernicosa - Nothofagus cunnunghamii*
AV-Ngum *Nothofagus gunnii*

Heathlands and related shrublands—temperate south-western Australia (A. S. George, A. J. M. Hopkins, and N. G. Marchant, *Ecosystems of the World,* Vol. 9A, pp. 211–230. Elsevier, Amsterdam, 1979; R. L. Specht and A. J. M. Hopkins, unpubl. data)

SW22 *Allocasuarina humilis - Calothamnus quadrifidus - Dryandra nivea*
SW46 *Allocasuarina humilis - Dryandra nivea - Hakea lissocarpha*
SW47b *Banksia attenuata - B. menziesii*
SW99a *Allocasuarina acutivalvis*

Tussock grasslands (R. L. Specht and M. P. Bolton, unpubl. data)

G8 *Astrebla elymoides - Iseilema membranaceum*
G11 *Dichanthium sericeum*
G18 *Astrebla* spp. - *Dichanthium fecundum - Iseilema* spp.
G19 *Astrebla elymoides - A. lappacea*
G56 *Xerochloa* spp. - *Eleocharis* spp. - *Oryza sativa*
G69 *Themeda triandra* (Basalt Plains, western Victoria)
G114 *Themeda triandra* (coastal headlands, Queensland)
GDanth *Danthonia caespitosa*

***Acacia* vegetation—subhumid, subtropical eastern Australia** (R. L. Specht, M. P. Bolton, and A. Specht, unpubl. data)

HW11 *Acacia harpophylla - Casuarina cristata*
DA8a *Acacia catenulata*
DA8b *Acacia shirleyi*
DA18 *Acacia cambagei*
DA40 *Acacia pendula - Atriplex nummularia*
DAargy *Acacia argyrodendron*
DAspar *Acacia sparsiflora - A. burrowii*
DAsuth *Acacia sutherlandii - Astrebla* spp.

***Acacia* vegetation—Australian Arid Zone** (R. L. Specht, M. P. Bolton, and A. Specht, unpubl. data)

DA14a *Acacia ligulata*
DA19 *Acacia aneura - Eucalyptus populnea - Eremophila mitchellii*
DA21 *Eucalyptus terminalis - Acacia ligulata - A dictyophleba*
DA22a *Acacia aneura - Senna* spp. - *Eremophila* spp.
DA22d *Acacia sclerosperma - A. aneura,* etc.
DA24 *Eucalyptus terminalis - Acacia dictyophleba - Senna artemisioides*
DAgeor *Acacia georginae*
DA41a *Acacia aneura - Alectryon oleifolius*
C24 *Acacia papyrocarpa - Maireana sedifolia*
DA46a *Acacia kempeana - A. aneura*
DA47 *Acacia acradenia - A. tenuissima - Senna artemisioides*
DA51 *Eucalyptus gamophylla - Acacia* spp. - *Dodonaea viscosa*
DA62 *Eucalyptus kingsmillii - E. youngiana - E. gamophylla - Triodia* sp.

DA62a *Eucalyptus gongylocarpa - Triodia basedowii*
DA63a *Eucalyptus gamophylla - Acacia* spp. *- Thryptomene maisonneuvei*
DA100g *Eucalyptus dichromopholoia - Acacia lysiphloia - Acacia* spp.
DA101c *Acacia pachycarpa - Eucalyptus dichromophloia*
Casdec *Allocasuarina decaisneana - Triodia basedowii*
DAstow *Acacia stowardii - Senna* spp. *- Eremophila* spp.
Erem *Eremophila* spp. *- Dodonaea* spp. *- Senna* spp. *- Acacia* spp.

Hummock grasslands—Australian Arid Zone (R. L. Specht and E. E. Hegarty, unpubl. data)

HG1 *Triodia pungens*
HG2 *Triodia basedowii*
HG3 *Triodia wiseana*
HG4 *Triodia intermedia - Eriachne obtusa*
HG5 *Triodia inutilis - Triodia* spp.
HG6 *Plectrachne schinzii*
HG7 *Triodia clelandii - T. irritans*
HG8 *Triodia basedowii - Zygochloa paradoxa*
HG9 *Triodia irritans*
HG10 *Triodia hubbardii*

Chenopod low shrublands—southern Australian Arid Zone (R. L. Specht and A. Specht, unpubl. data)

C4 *Atriplex vesicaria - Maireana sedifolia*
C7 *Myoporum platycarpum - Atriplex vesicaria* ± *Cratystylis* sp.
C13 *Alectryon oleifolius* ± *Atriplex vesicaria* ± *Maireana* spp.
C20 *Halosarcia* spp. (inland salt marsh)
C21 *Atriplex vesicaria - Maireana aphylla - M. planifolia*
C22 *Casuarina pauper - Maireana pyramidata*
C23a *Atriplex vesicaria - A. nummularia - Acacia pendula*
C24 *Acacia papyrocarpa - Maireana sedifolia*

Aquatic vegetation—tropical and subtropical northern Australia (R. L. Specht and M. P. Bolton, unpubl. data)

A12 *Monochoria cyanea - Potamogeton tricarinatus*
A15/26 *Caldesia oligococca - Eleocharis dulcis - Nymphaea gigantea*
A27 *Caldesia oligococca - Nymphaea gigantea*
A28 *Typha domingensis - Caldesia oligococca*
A29 *Typha domingensis*
Restio *Leptocarpus spathaceus - Restio tetraphyllus*

Aquatic vegetation—temperate southern Australia (H. I. Aston, *Aquatic Plants of Australia,* Melbourne University Press, Melbourne, 1973; R. L. Specht and M. P. Bolton, unpubl. data)

A16 *Eleocharis acuta - Isolepis fluitans - Lepilaena* spp.
A17 *Lemna minor - Isoetes muelleri*
A18/19 *Eleocharis acuta - E. sphacelata*
A22 *Eleocharis acuta - E. sphacelata - Typha* spp. *- Maundia* sp.
A40/43 *Eleocharis acuta - E. sphacelata - Typha* spp. *- Marsilea* sp.
A41 *Eleocharis acuta - E. sphacelata - Typha* spp. *- Myriophyllum* sp.
A42 *Eleocharis acuta - E. sphacelata - Typha* spp.
A46 *Eleocharis acuta - Typha* spp. *- Potamogeton crispus*
A47 *Eleocharis acuta - Typha* spp.
Baumea *Baumea juncea - B. rubiginosa*
Gahnia *Gahnia filum - G. trifida*

Coastal dune vegetation (R. L. Specht, *Ecosystems of the World,* Vol. 2B, pp. 223–237, 1993; Vol. 2C, pp. 483–495, 1997, Elsevier, Amsterdam)

CD8 *Olearia axillaris - Capparis spinosa - Spinifex longifolius*
CD12 *Banksia integrifolia - Casuarina equisetifolia - Spinifex sericeus*
CD13 *Calophyllum inophyllum - Spinifex sericeus*
CD14 *Calophyllum inophyllum - Casuarina equisetifolia - Spinifex longifolius*
CD15 *Casuarina equisetifolia - Argusia argentea - Spinifex longifolius*
CD18 *Olearia axillaris - Acacia cyclops - Spinifex hirsutus*
CD19 *Allocasuarina verticillata - Olearia axillaris - Spinifex sericeus*
CD20 *Banksia integrifolia - Leptospermum laevigatum - Leucopogon* sp.
CD21 *Banksia integrifolia - Leptospermum laevigatum - Myoporum* sp.
CD22 *Banksia integrifolia - Leptospermum laevigatum - Cupaniopsis* sp.
CD23 *Banksia integrifolia - Casuarina equisetifolia - Isolepis nodosa*
Pisonia *Pisonia grandis*

Coastal wetland vegetation (mangroves, salt marshes, and brackish wetlands) (P. Saenger, R. L. Specht, *et al.*, *Ecosystems of the World,* Vol. 1, pp. 293–345, Elsevier, Amsterdam, 1977; R. L. Specht, *Ecological Biogeography of Australia,* pp. 575–590, Dr. W. Junk, The Hague, 1981)

CW5 *Avicennia marina - Aegiceras corniculatum - Halosarcia indica*
CW5d/19d *Casuarina glauca*
CW6 *Rhizophora stylosa - Avicennia marina - Aegialitis annulata*
Nypa *Nypa fruticans*
Melacac *Melaleuca acacioides*

CW7 *Aegialitis annulata - Avicennia marina - Rhizophora* spp.
CW19 *Avicennia marina - Aegiceras corniculatum - Halosarcia pergranulata*
CW36 *Avicennia marina - Halosarcia halocnemoides*
Melhal *Melaleuca halmaturorum*
CW37 *Avicennia marina - Halosarcia pergranulata*
Meleric *Melaleuca ericifolia - M. squarrosa*
CW8 *Sclerostegia arbuscula* (salt marsh)

V. BIOGEOGRAPHIC REGIONS IN AUSTRALIA

The presence or absence of TWINSPAN Floristic Groups in each 1° latitude by 1° longitude grid-cell of the Australian continent was analyzed using the PATN micro-computer Pattern Analysis program. The symmetric form of the Kulczynski measure with $\beta \equiv -0.1$ was used to determine the association between community groups in each grid-cell. This has been found to produce the most useful classification for such ecological applications (D. P. Faith, P. R. Minchin, and L. Belbin, *Vegetatio* **69**, 57–68, 1987) and in this case produced the most "sensible" results. The results are plotted as a dendrogram in Fig. 2. The resulting distribution of the biogeographic regions identified by the foregoing analysis is plotted in Fig. 3.

The first subdivision of the Kulczynski Symmetric dendrogram (see Fig. 2) separated the arid and semiarid biogeographic regions of Western Australia, Northern Territory, South Australia, and north-western Victoria. Three biogeographic regions (AN2–4) were defined in the northern section of the arid zone, an area that receives mostly summer rainfall. The biogeographic regions of Central Australia (AC) and the Ashburton (ANI) of Western Australia receive both summer and winter rainfall. Although summer rains occur erratically, winter rainfall is predominant in the southern arid zone (A1a–b, A2a–b, A3, A4a–b) and in the semiarid climatic zone (SWI1–2 and S1–3). The TWINSPAN Floristic Groups on the mid-Tertiary limestone sediments from the Nullarbor Plain (A3) across South Australia (S1 and S2) into north-western Victoria (S3) are closely related.

The complex of TWINSPAN Floristic Groups (A1a–b) in the southern section of the arid zone of Western Australia are closely interrelated (and probably should not be separated); they are affiliated with the vegetation of the Ashburton region (AN1), where summer rainfall predominates. The northern section (A2a–b) of the arid zone in South Australia is distinctly separated from the southern arid vegetation (A4a–b) in which outliers of the mallee open-scrub (dominant in the semi-arid climate of S1–3) are to be found.

The second subdivision of the Kulczynski Symmetric dendrogram differentiated both the vegetation of the humid/perhumid section of south-western Western Australia (SW1–2, with a Mediterranean-type climate) and all the biogeographic regions in eastern Australia (where rainfall is evenly distributed throughout the year). Biogeographic regions were defined in the humid/perhumid climates of Tasmania to south-eastern Queensland (T, Tasmania; SV, Southern Victoria; and E1–3, Eastern Australia). Subhumid biogeographic regions were defined in central Queensland (CQ1–3) and in central New South Wales (CN), with related biogeographic regions (BT, the Barkly Tableland; CC, the Channel Country; and A5, Far Western New South Wales) in the semi-arid to arid zones of both states.

The third subdivision of the Kulczynski Symmetric dendrogram defined the biogeographic regions in northern Australia, where summer rainfall alternates with winter drought. The biogeographic regions of Cape York Peninsula (NE1) and the Gulf of Carpentaria (NE2) in Queensland are closely related to the wet tropical region (NE3) that extends southwards along the coast toward the tropic of Capricorn.

The TWINSPAN Floristic Groups of the humid/subhumid section of the Kimberley (N3) and the Top End of the Northern Territory (N1–2) are interrelated and were differentiated from the inland, semi-arid/subhumid biogeographic regions (NI1–3) across northern Australia and the Hamersley Ranges (HR) in Western Australia.

VI. DIAGNOSTIC FLORISTIC GROUPS IN AUSTRALIAN BIOGEOGRAPHIC REGIONS

Australian biogeographic regions (shown by alphanumeric codes in Figs. 2 and 3) were defined using the classificatory program PATN on the basis of the spatial distribution of the TWINSPAN floristic groups. The PATN regions and their associated TWINSPAN groups are shown below.

N1 (humid north-western W.A.)—RF34(69), MRF12a, T371, T372
N2 (humid N.T.)—MRF12a, T187a, T371, T372, T373a, G56, H7
N3 (humid Cape York Peninsula and Gulf, Qld.)—RF35(140), RF35(141), T370, T372, T373a, G18, H12, H13
N4 (Broome-Derby, W.A.)—RF16(132), T372

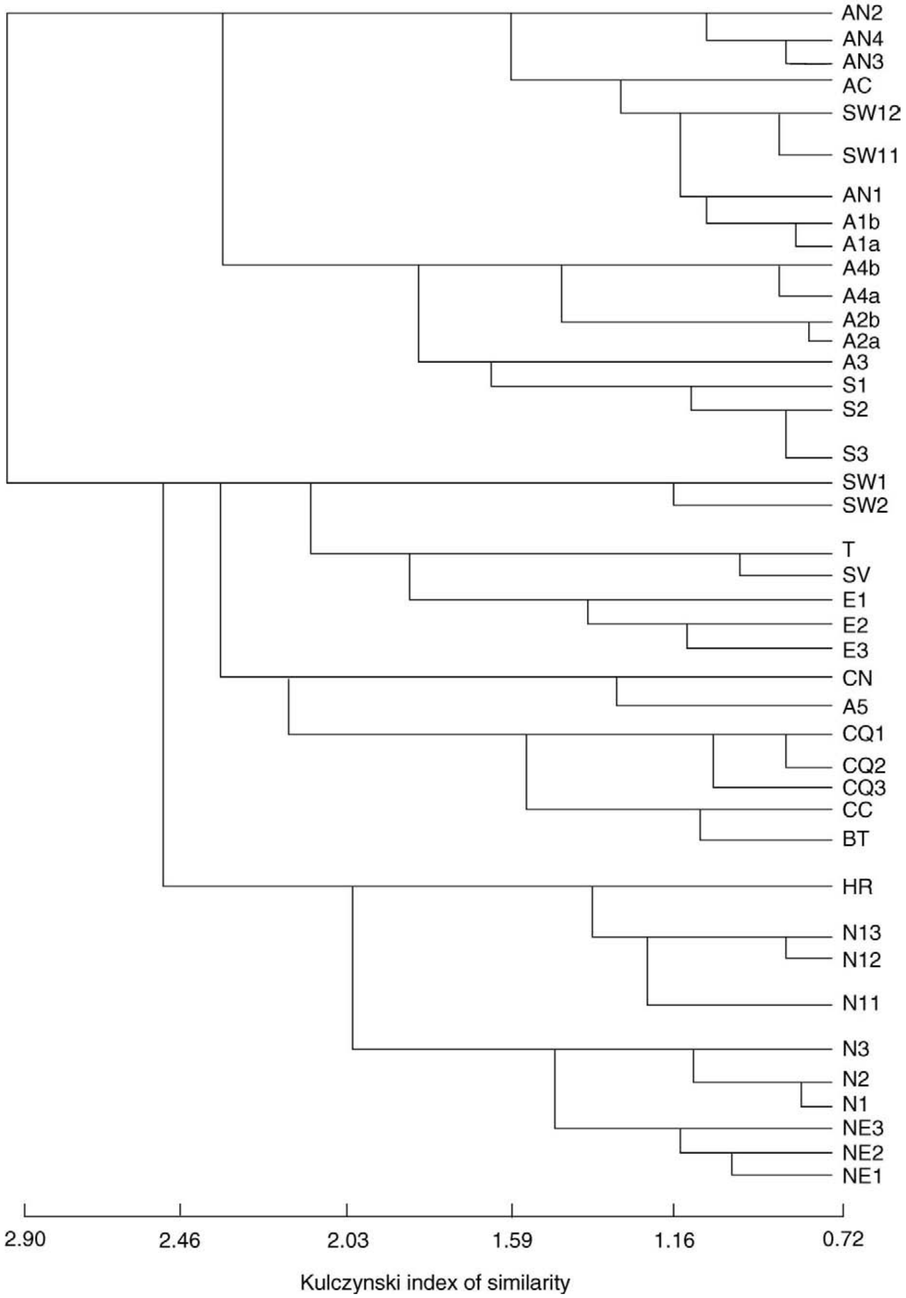


FIGURE 2 Dendrogram showing the relationships of the biogeographic regions of Australia derived from a Kulczynski Symmetric Pattern Analysis of the major TWINSPAN Floristic Groups recorded in 1° latitude by 1° longitude grid-cells throughout Australia. (From Specht *et al.*, 1995; Specht and Specht, 1999.)

N5 (southern Kimberley region, W.A.)—RF16(132), T187a, T371, T372
N6 (subhumid N.T.)—T187a, T371, T372, H7
N7 (semi-arid N.T.)—T187a, T372, DA8b
N8 (Pilbara region, W.A.)—T372, DA100g
NE1 (subhumid Cape York Peninsula, Qld.)—T188, T189, T368, T369, T371, T714, T801, G18
NE2 (perhumid north-eastern Qld.)—RF35(142), RF35(143), RF36(144), RF36(145), T368, T369, T371, T710, T712, T713, T714, T715, T801

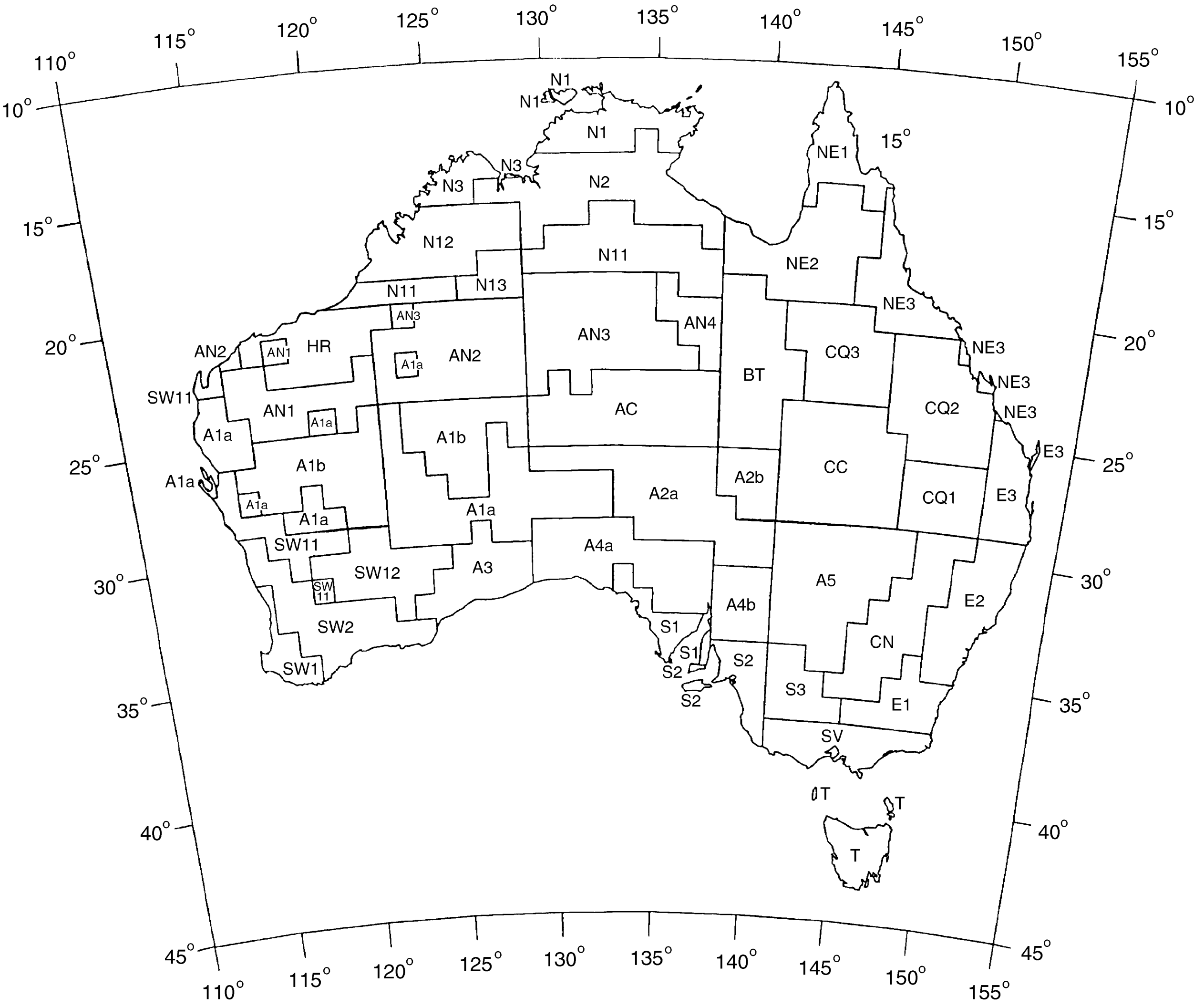


FIGURE 3 Biogeographic regions of Australia derived from a Kulczynski Symmetric Pattern Analysis of the major TWINSPAN Floristic Groups recorded in 1° latitude by 1° longitude grid-cells throughout Australia. (From Specht *et al.*, 1995; Specht and Specht, 1999.)

NE3 (subhumid north-central highlands, Qld.)—T188, T189, T368, T369, T714, DA8b

NE4 (subhumid central Qld.)—RF46(184), T188, T364, T352b, T365, T369, T714, T716a, T717, HW11, DA8a

NE5 (subhumid south-eastern Qld.)—T188, T360, T363a, T364, T365, T369, T716a, T717, Calcolu, HW11

BT (Barkly Tableland, Qld.)—T188, G8, G19

CC (Channel Country, Qld.)—T188, DA18, DA19, C20

SE1 (humid south-eastern Qld.)—RF45(180), RF41(332), T354, T707, T710, T711, T713, T714, T715, T716a, T717, G11, H82, H166, H167

SE2 (perhumid Border Ranges, N.S.W.–Qld.)—RF42(169), RF40(322), RF40(325), T346, T707, T710, H82, H84, H166, H167

SE3 (central to northern coasts and highlands, N.S.W.)—T334, T347, T350, T354, T707, T710, T716b, H82, H84, H166, H167

SE4 (south-eastern and montane N.S.W.–Vic.)—RF25, T38, T79, T156, T157, T334, T335, T347, T348, T349, T350, AV4, AV6, H84

SE5 (south-western to south-eastern Vic.)—RF24, T331, Eucpryor, T341, T342, T666, T667, G69, H170, H171

SE6 (Tasmania)—RF7, RF13, T8, T36, T40, T148, T149, Eucnit, T666, T667, AV14, AV15, H170, H171

SE7 (northern tablelands, N.S.W.)—T338, T339, T362, T363a, T717, Calcolu

SE8 (southern and central tablelands, N.S.W.)—T190, T338, T339, T341, T362, T363a, T716b, Calcolu
SE9 (humid Mt. Lofty Ranges and South East District, S.A.)—T190, T328a, T329, T330, T331, T667, H170
S1 (Nullarbor Plain, S.A.)—M13, M45, C4, C20
S2 (northern Eyre and Yorke Peninsulas, S.A.)—M25, M45, C24, DA14a, H44b
S3 (subhumid S.A.)—T190, T328a, T329, T330, T331, T366a, T366b, T667, M13, M25, H44b, H170
S4 (north-western Vic.)—T190, T329, T362, T366a, T366b, M10, M25, M45, H44a, H44b
SW1 (humid to perhumid south-western W.A.)—SW4a, SW20, SW43a, SW47, SW84, SW85
SW2 (subhumid to semi-arid south-western W.A.)—SW22, SW46, SW96a, SW97a, SW100, SW196
A1a (south-eastern Pilbara District, W.A.)—C20, DA14a, DA22a, DA51, DA62, DA63a
A1b (southern part of Murchison District to Great Victoria Desert, W.A.)—C20, DA22a, DA62, SW197
A1c (northern part of Murchison District to Great Victoria Desert, W.A.)—C20, DA22a, DA62, DA63a
A1d (southern N.T.)—C20, DA21, DA46a
A1e (Great Victoria Desert, S.A.)—C20, C22, DA62
A2a (Great Sandy Desert, W.A.)—C20, DA100g, DA101c, HG6
A2b (Tanami Desert, N.T.)—C20, DA47, HG6
A3a (northern S.A.)—C20, DA21, DA41a, DA46a
A3b (Simpson Desert, S.A.)—C20, HG2, HG8
A4a1 (Exmouth Gulf, W.A.)—C20, DA14a
A4a2 (north-eastern Wheat Belt, W.A.)—C20, DA14a, SW100, SW196
A4b (Goldfields, W.A.)—C20, DA14a, DA62, SW26, SW51, SW99a, SW101b
A5a (Nullarbor Plain to Flinders Ranges, S.A.)—C4, C7, C13, C20, C21, C24, DA41a
A5b (North East District, S.A.)—C4, C7, C13, C20, C22, DA41a, T366a
A6 (western N.S.W.)—C4, C20, C22, C23a, DA41a, T366a

VII. SPECIES RICHNESS OF AUSTRALIAN ECOSYSTEMS

The potential for annual foliage growth increases in the 10°C temperature gradient from temperate to tropical Australia (Table II). The numbers of both stems and species recorded in the overstory (per hectare) increase exponentially (Fig. 4a; after Specht and Specht, 1993, 1994).

TABLE II

Structural Characteristics of Mature Australian Plant Communities[a]

Climate	Stand height (m)	Stand density (stems ha^{-1})	Foliage projective cover (%)	Annual shoot growth (kg ha^{-1})	Overstory (spp. ha^{-1})	Understory (spp. ha^{-1})
Tropical						
Perhumid	30–(45[b])–24	>5000–2500	100–65	12,500	140	150[c]
Humid	24–17	2500–800	65–45	7600	35	50
Subhumid	17–14	800–450	45–35	5700	20	30
Semiarid	14–10	450–200	35–25	4800	12	25
Arid	<10	<200	<25	3600	8	15
Subtropical						
Perhumid	30–(50[b])–22	4500–1200	100–65	10,500	90	120[c]
Humid	22–15	1200–500	65–45	6000	20	60
Subhumid	15–12	500–300	45–35	4400	10	40
Semiarid	12–8	300–200	35–25	3500	6	35
Arid	<8	<200	<25	2800	4	25
Temperate						
Perhumid	30–(60[b])–20	550–400	100–65	4800	10	45[c]
Humid	20–14	400–300	65–45	2700	5	45[c]
Subhumid	14–11	300–250	45–35	1600	4	44
Semiarid	11–7	250–200	35–25	1100	3	43
Arid	<7	<200	<25	600	2	40

[a] From R. L. Specht and A. Specht (1989a) and A. Specht and R. L. Specht (1993, 1994).

[b] Stand heights of emergent eucalypts, with high nitrate-reductase activity, at the edge of rainforest stands.

[c] The species richness of understories in perhumid to humid climates contains a mixture of regenerating overstory species as well as shade-loving vascular plants and cryptogams.

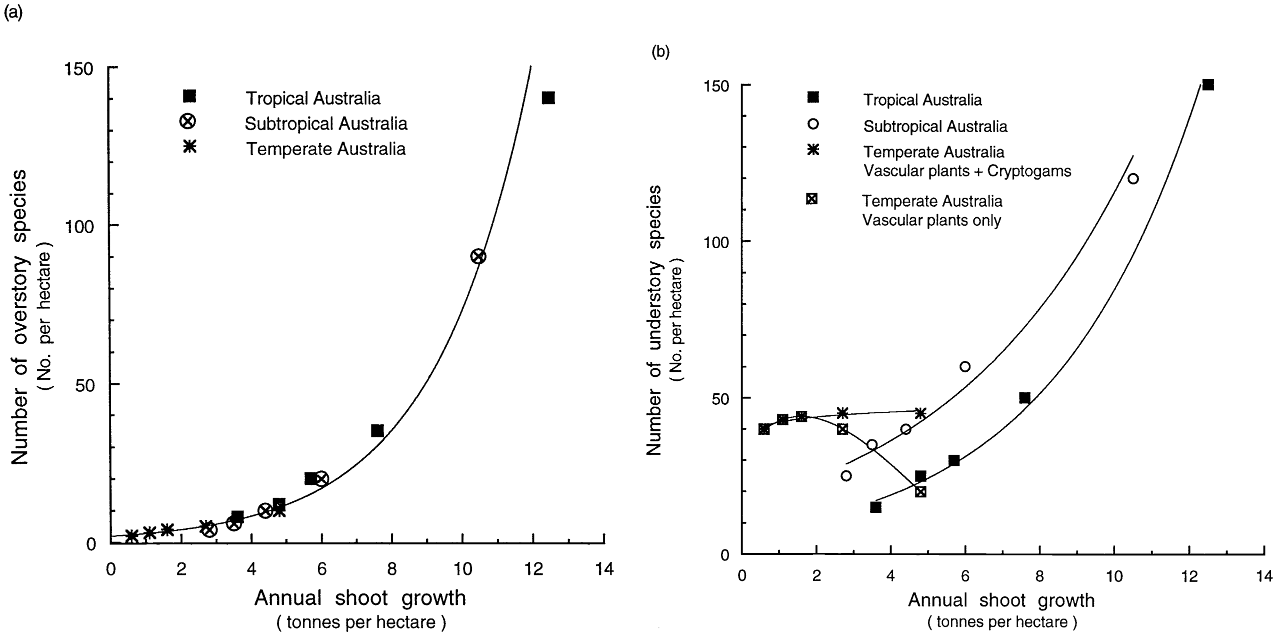


FIGURE 4 (a) Species richness (number of species per hectare) of the overstory increases exponentially as annual shoot growth of this stratum increases. Species richness (N) increases exponentially (1) in the evaporative gradient from the arid zone (Evaporative Coefficient $k < 0.35 \times 10^{-2}$) to the perhumid zone (Evaporative Coefficient $k = 1.0 \times 10^{-2}$), and (2) in the 10°C temperature gradient from temperate to subtropical to tropical Australia. (From Specht and Specht, 1993.) Warm temperate ln Overstory $N = 0.21 + 2.05\text{-}k$ ($n = 16$, $r^2 = 0.71$); Subtropical ln Overstory $N = 0.01 + 4.95\text{-}k$ ($n = 19$, $r^2 = 0.93$); Tropical ln Overstory $N = 0.92 + 4.00\text{-}k$ ($n = 29$, $r^2 = 0.98$). (b) The low amount of solar radiation that penetrates the overstory to reach the understory in the temperate latitudes of Australia has little effect on species richness (number of species per hectare) of this stratum. In the tropics and subtropics, the relationship of species richness of the understory to annual shoot growth (of the upper stratum) is similar to that shown by the overstory. (After Specht and Specht, 1993; 1994; 1999)

Photosynthesis in overstory leaves, and hence potential annual foliage growth, depends primarily on the amount of solar radiation incident on the plant community, a level that declines from tropical to temperate Australia (Specht and Specht, 1999). In the open-structured plant communities of the wet–dry tropics and subtropics, a high percentage of direct-beam solar radiation reaches the understory through the gaps in the overstory canopy; species richness (the number of species per hectare) of the understory is higher than that of the exposed overstory (Fig. 4b), and follows the exponential curve shown for the overstory in Fig. 4a. In more temperate latitudes, solar radiation reaches the Earth at a lower angle, and consequently the amount of radiation that reaches the understory is low (Specht *et al.*, 1992; Specht and Specht, 1999). Species richness of the understory in temperate latitudes increases only slightly from the semi-arid to the perhumid zone, with the proportion of cryptogamic species rising strongly in the most humid climates.

On nutrient-poor soils (residual lateritic soils, arenaceous, acid granitic, and serpentinitic soils, Quaternary sands, etc.), common in many parts of Australia, the establishment of a symbiotic association of mycorrhizae with seedling rootlets of Myrtaceae is reduced as available soil water decreases (J. P. Burrell, *Aust. J. Bot.* **29**, 747–764, 1981), thus affecting the density of stems in the overstory. As soil depth decreases, less soil water is available for evapotranspiration, the transport of phosphate and other nutrient ions in the transpiration stream is reduced, and the growth of the overstory is limited (Batianoff *et al.*, 1997)—thus the foliage projective cover of the overstory, the annual shoot growth (per hectare), and the stand height are all reduced (Fig. 5a). As more solar radiation penetrates the gaps between the trees due to the reduction in overstory growth, the foliage projective cover of the understory increases; the sum of the foliage covers of overstory and understory strata remains a constant in equilibrium with the evaporative aerodynamics (Fig. 1) of the area (Fig. 5b; Specht and Morgan, 1981; Specht, 1983; Specht *et al.*, 1990). The species richness of the overstory on nutrient-poor soils is low, below that on higher-fertility soils in the same evaporative regime (Specht and Specht, 1989b), whereas the species richness of the understory increases

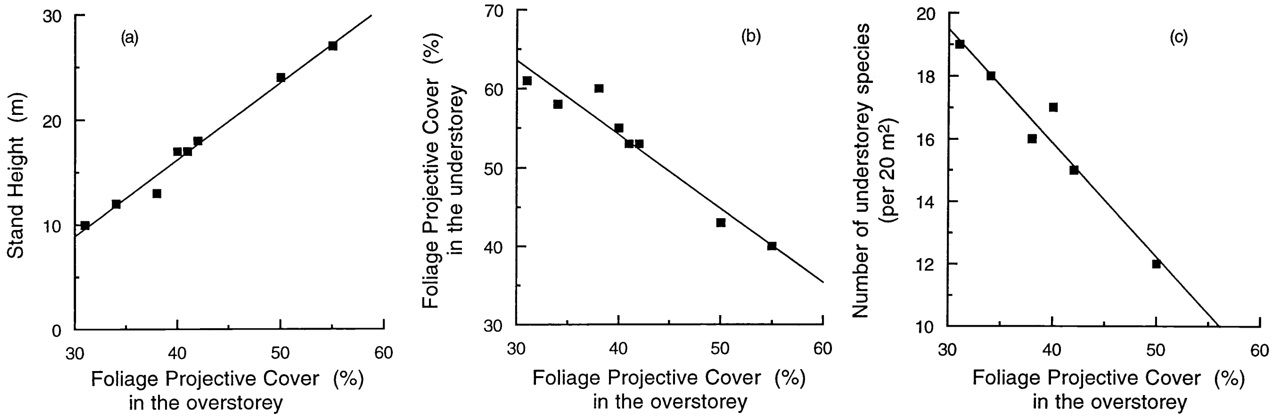


FIGURE 5 The relationship between foliage cover of the overstory and (a) stand height of the overstory; (b) foliage cover of the understory; and (c) species richness of the understory—in the open-forest on the serpentinite soil catena in Central Queensland. (After Batianoff *et al.*, 1997)

as more solar radiation penetrates the overstory (Fig. 5c).

The shading effect of the overstory on the species richness in the understory has been demonstrated when long-lived shrubby species, such as *Banksia* (Proteaceae), regenerate from seed or resprouts in a secondary succession following the fire that regularly razes Australian heathlands (Fig. 6). Application of phosphate fertilizer to these nutrient-poor ecosystems results in more rapid growth of overstory shrubs to the detriment of the understory species (Specht and Specht, 1989c).

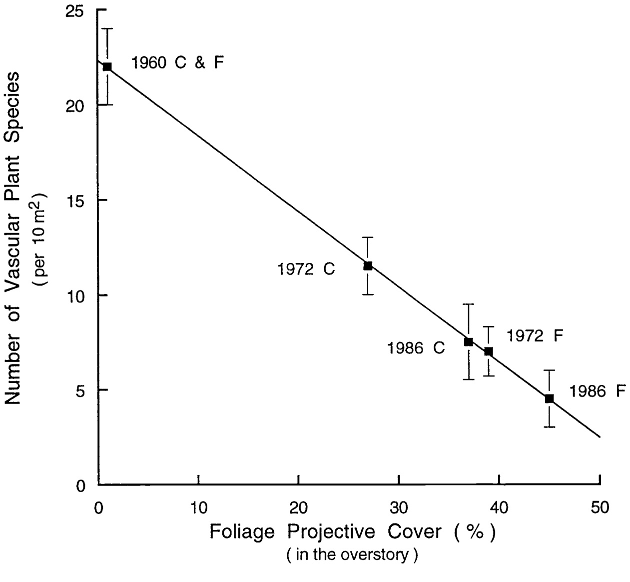


FIGURE 6 The species richness of Dark Island heathland, South Australia (Specht and Specht, 1989b) decreased as foliage cover of the overstory (*Banksia ornata, B. marginata, Allocasuarina pusilla,* and *Xanthorrhoea caespitosa*) increased with time after fire (control C) and with phosphate fertilizer (F).

The species richness of vascular plants in both overstory and understory strata is related to the amount of solar energy fixed as annual shoot growth. The community-physiological processes that produce these predictable values of alpha diversity need urgent scientific investigation. The species richness of vertebrates recorded in a plant community parallels the number of vascular plants found in the understory stratum (Fig. 7) and thus must also be related to the potential energy fixation (Braithwaite *et al.,* 1985; P. C. Catling, in Specht, 1988; Cody, 1994a, 1994b; Specht and Specht, 1999). It would also appear from the limited data available that the species richness of epigaeic invertebrates collected in pit-fall traps decreases as the annual input of leaf litter from the overstory decreases from the humid to the semi-arid climatic zone in the Mediterranean-type climate of southern Australia (Table III; after Greenslade and Majer, 1985, and in Specht, 1988).

VIII. COMMUNITY DIVERSITY OF AUSTRALIAN ECOSYSTEMS

Over large areas, such as a 1° latitude by 1° longitude grid-cell, the diversity of habitats within each evaporative climate is high, resulting in a high diversity of TWINSPAN Floristic Groups, each defined by diagnostic species, within the floristic continuum. The number of Floristic Groups within each large grid-cell is corre-

TABLE III

Number of Groups[a] of Epigaeic Invertebrates Collected (Throughout the Year) in Pit-fall Traps Established within Four Vegetation Formations Typical of Humid to Semiarid Mediterranean-Type Climates of Southern Australia[b]

Locality	Vegetation formation	Evaporative Coefficient	Annual leaf litter (kg ha^{-1})	Epigaeic invertebrates (no. groups)
Dwellingup, W.A. (32°43′S, 116°04′E)	Eucalypt open-forest	0.53×10^{-2}	1300	19
Kuitpo, S.A. (33°15′S, 138°43′E)	Eucalypt woodland	0.45×10^{-2}	1050	13
Wyperfield, Vic. (35°35′S, 142°00′E)	Eucalypt open-scrub	0.42×10^{-2}	950	12
Wyperfield, Vic. (35°35′S, 142°00′E)	Heathland	0.40×10^{-2}	850	10

[a] Invertebrate groups—Acarina, Araneae, Isopoda, Amphipoda, Diplopoda, Chilopoda, Collembola, Insecta (Thysanura, Blattodea, Isoptera, Dermaptera, Orthoptera, Embioptera, Psocoptera, Hemiptera, Homoptera, Thysanoptera, Coleoptera, Diptera, Lepidoptera, and Hymenoptera including Formicidae).

[b] After Greenslade and Majer (1985) and Greenslade and Majer, in Specht (1988).

[c] Monthly actual/potential evapotranspiration per millimeter of available water.

lated with the optimal annual shoot growth for the evaporative climate of the area (Fig. 8a). The number of vertebrate species also appears to be correlated with the optimal shoot growth of plant communities in the grid-cell (Fig. 8b).

As foliage growth is reduced by the dearth of phosphate ions available in the nutrient-poor lateritic soils of northern, central, and western Australia [also in the Fleurieu Peninsula (35°34′S, 138°20′E) and on Kangaroo Island (35°50′S, 137°15′E) of South Australia and in the Brisbane Ranges (37°55′S, 144°20′E) of Victoria], stem density and species richness of the overstory trees

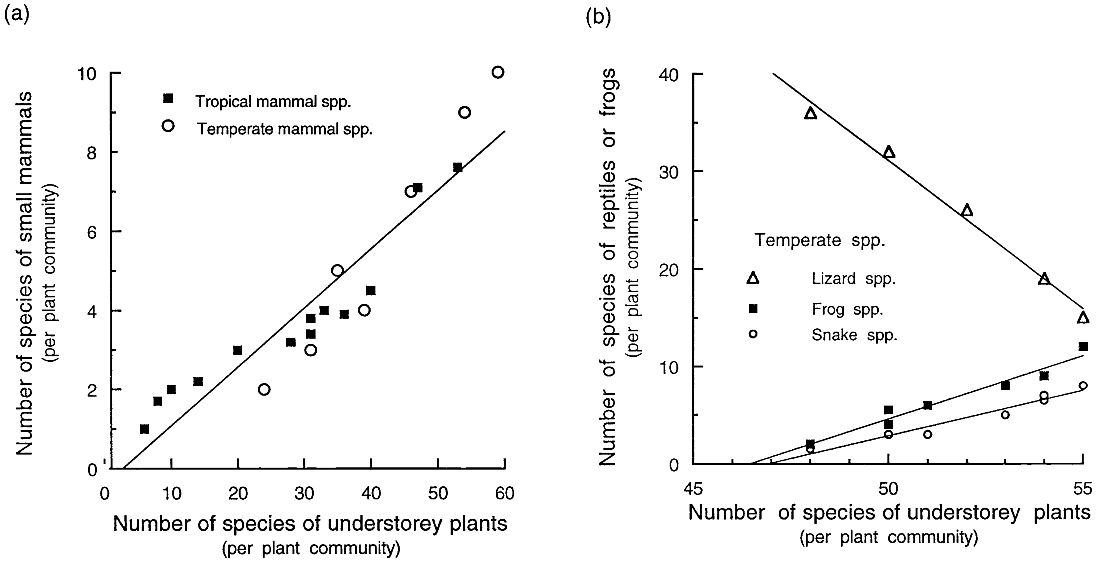


FIGURE 7 (a) The number of species of small mammals recorded in ecosystems in tropical and temperate Australia plotted against the number of species of vascular plants recorded in the understory of each plant community. (b) The number of species of lizards, snakes, and frogs recorded in ecosystems in temperate Australia plotted against the number of species of vascular plants recorded in the understory of each plant community. (After Specht and Specht, 1999; Braithwaite *et al.*, 1985 and R. C. Catling in Specht, 1988)

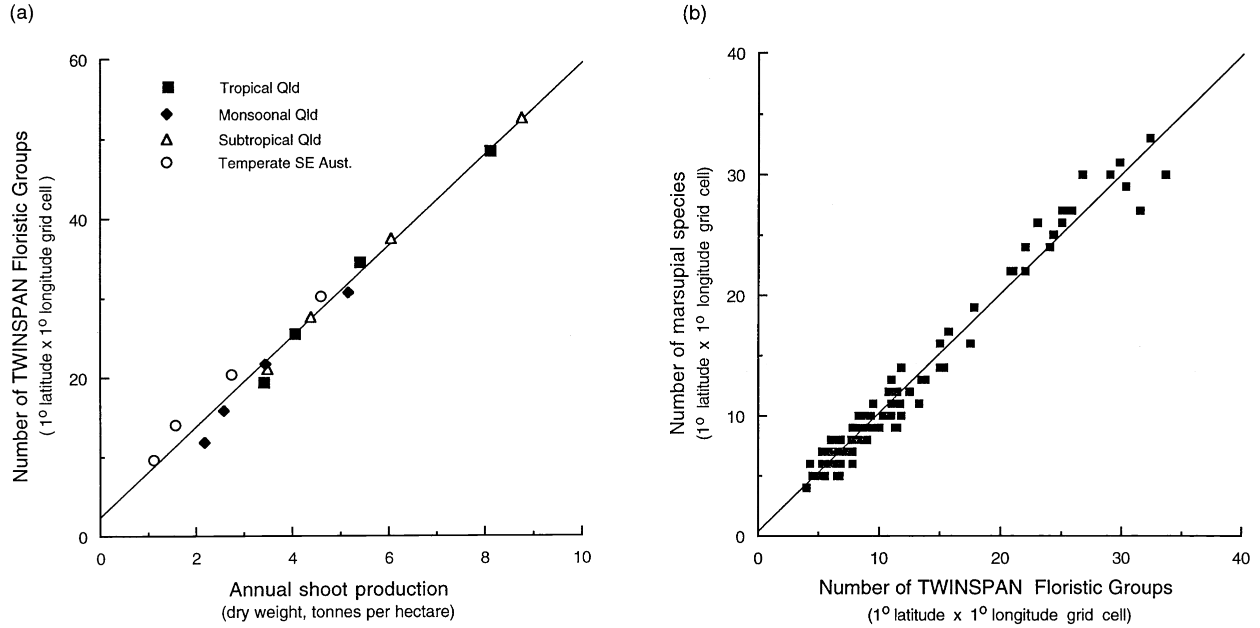


FIGURE 8 (a) Relationship between the number of TWINSPAN Floristic Groups (per 1° latitude by 1° longitude grid-cell) and potential annual shoot growth (per hectare) in the same evaporative climate. (From Specht and Specht, 1999.). (b) Number of marsupial species plotted against the number of TWINSPAN Floristic Groups found in 1° latitude by 1° longitude grid-cells. (From Specht and Specht, 1999, after Pianka and Schall, 1981.)

are lower in these plant communities. The number of TWINSPAN Floristic Groups is therefore lower on these soils of low fertility (Fig. 9).

IX. CONCLUSIONS

The vegetation of Australia in the Cretaceous–Early Tertiary, fifty to a hundred million years ago, had developed on infertile lateritic soils which extended across the Gondwanan super-continent of which Australia was part. The tropical and temperate rainforests, savanna and healthy floras of Australia therefore show strong family and generic (sometimes specific) affinities with the floras of India, Africa and, to a lesser extent, with South America (Specht, 1981).

Although Australia was at this time contiguous with Antarctica at latitude 60–65° S, palaeo-oxygen analyses of sea-bed sediments indicate that the climate was subtropical with a mean annual temperature of 19.5°C (Specht *et al.*, 1992). Increasing aridity during the Tertiary, as well as the movement of the Australian Tectonic Plate northward towards the equator at 66 mm per year, imposed great stresses on the Australian vegetation. Remnants of the Gondwanan flora persisted as fragments in more humid habitats. In addition, a great

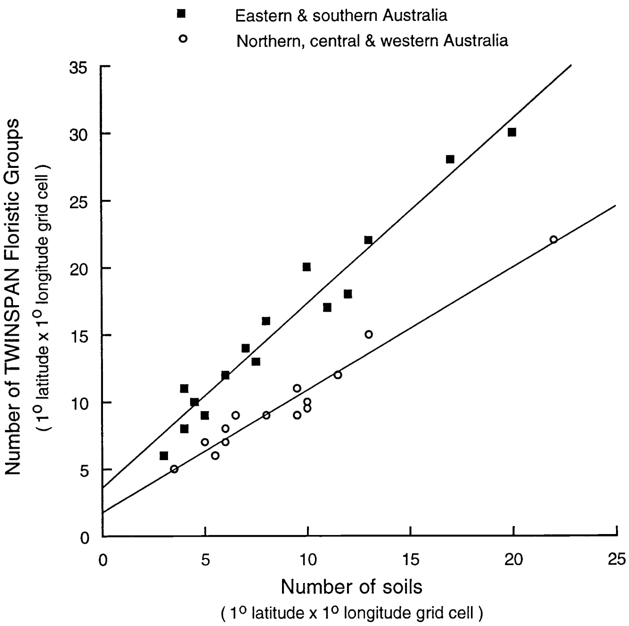


FIGURE 9 The number of TWINSPAN Floristic Groups plotted against the number of principal soil profiles (K. H. Northcote *et al.*, 1960–1968) found in the same grid-cell of 1° latitude by 1° longitude dimensions in eastern and southern Australia and on the nutrient-poor soils of northern, central, and western Australia. (After Specht and Specht, 1999 and unpubl.)

diversity of taxa evolved during the Early Tertiary to occupy the greater part of the continent which today experiences varying degrees of seasonal drought (Specht, 1981; Specht and Specht, 1999). The convergence of the Australian Plate with the Sunda Plate about 15 million years ago allowed some migration of the Australian flora northward onto Sundaland, but little, if any, migration southward from southeast Asia.

The evaporative and frictional forces of the atmosphere, flowing over and through the vegetation across the Australian continent, moulded foliage distribution and structure within all plant communities; distinctive plant formations resulted from the perhumid to the arid evaporative zone, from the tropical to the temperate region. Species richness (per hectare) of vascular plants in the overstory of Australian plant communities is correlated with the annual shoot growth (per hectare) of that stratum. Similarly, the species richness of vascular plants in the understory is related to annual shoot growth in that stratum—an attribute influenced by the amount of solar radiation transmitted through the overstory stratum.

The species richness (per hectare) of non-arboreal vertebrates is correlated with species richness of the understory stratum. Species richness of epigaeic invertebrates appears to depend on the annual input of leaf litter from the overstory foliage. The diversity of TWINSPAN Floristic Groups and associated vertebrates is also correlated with the optimal annual shoot growth (per hectare) of the overstory stratum of Australian plant communities.

The community-physiological processes that determine the species richness of vascular plants, vertebrates, and invertebrates and the diversity of Floristic Groups need urgent scientific investigation before they will be better understood.

See Also the Following Articles

AFRICA, ECOSYSTEMS OF • ANTARCTIC ECOSYSTEMS • ASIA, ECOSYSTEMS OF • SOUTH AMERICA, ECOSYSTEMS OF

Bibliography

Batianoff, G. N., Reeves, R. D., and Specht, R. L. (1997). The effect of serpentine on vegetation structure, species diversity and endemism in Central Queensland. In *The Ecology of Ultramafic and Metalliferous Areas* (T. Jaffré, R. D. Reeves, and T. Becquer, eds.), pp. 147–153. ORSTOM Centre de Nouméa, New Caledonia.

Braithwaite, R. W., Winter, J. W., Taylor, J. A., and Parker, B. S. (1985). Patterns of diversity and structure of mammalian assemblages in the Australian tropics. *Aust. Mammal.* **8**, 171–186.

Cody, M. L. (1994a). Bird diversity within and among Australian heathlands. In *Plant–Animal Interactions in Mediterranean-Type Ecosystems.* (M. Arianoutsou and R. H. Groves, eds.), pp. 47–61. Kluwer Academic Publishers, Dordrecht, Netherlands.

Cody, M. L. (1994b). Mulga bird communities. I. Species composition and predictability across Australia. *Aust. J. Ecol.* **19**, 206–219.

Greenslade, P., and Majer, J. D. (eds.). (1985). *Soil and Litter Invertebrates of Australian Mediterranean-Type Ecosystems,* Bulletin No. 12. Western Australian Institute of Technology, School of Biology, Perth.

Northcote, K. H., Hubble, G. D., Isbell, R. F., Nicolls, K. D., Thompson, C. H. (1960–1968). *Atlas of Australian Soils.* Sheets 1–10. CSIRO (Aust.) and Melbourne University Press, Melbourne.

Pianka, E. R., and Schall, J. J. (1981). Species densities of Australian vertebrates. In *Ecological Biogeography of Australia.* (A. Keast, ed.), pp. 1675–1694. Dr. W. Junk, The Hague, Netherlands.

Specht, A., and Specht, R. L. (1993). Species richness and canopy productivity of Australian plant communities. *Biodiversity & Conservation* **2**, 152–167.

Specht, A., and Specht, R. L. (1994). Biodiversity of overstory trees in relation to canopy productivity and stand density in the climatic gradient from warm temperate to tropical Australia. *Biodiversity Lett.* **2**, 39–45.

Specht, R. L. (1972). Water use by perennial, evergreen plant communities in Australia and Papua New Guinea. *Aust. J. Bot.* **20**, 273–299.

Specht, R. L. (1981). Evolution of the Australian flora: Some generalizations. In *Ecological Biogeography of Australia* (A. Keast, ed.), pp. 783–806. Dr. W. Junk, The Hague, Netherlands.

Specht, R. L. (1983). Foliage projective covers of overstory and understory strata of mature vegetation in Australia. *Aust. J. Ecol.* **8**, 433–439.

Specht, R. L. (ed.). (1988). *Mediterranean-Type Ecosystems. A Data Source Book.* Kluwer Academic Publishers, Dordrecht, Netherlands.

Specht, R. L., and Morgan, D. G. (1981). The balance between the foliage projective covers of overstory and understory strata in Australian vegetation. *Aust. J. Ecol.* **6**, 193–202.

Specht, R. L., and Specht, A. (1989a). Canopy structure in *Eucalyptus*-dominated communities in Australia along climatic gradients. *Oecol. Plant.* **10**, 191–202.

Specht, R. L., and Specht, A. (1989b). Species richness of overstory strata in Australian plant communities. The influence of overstory growth rates. *Aust. J. Bot.* **37**, 321–336.

Specht, R. L., and Specht, A. (1989c). Species richness of sclerophyll (heathy) plant communities in Australia. The influence of overstory cover. *Aust. J. Bot.* **37**, 337–350.

Specht, R. L., and Specht, A. (1999). *Australian Plant Communities: Dynamics of Structure, Growth and Biodiversity*. Oxford University Press, Melbourne.

Specht, R. L., Grundy, R. I., and Specht, A. (†R. Berliner) (1990). Species richness of plant communities: Relationship with community growth and structure. *Israel J. Bot.* **39**, 465–480.

Specht, R. L., Dettmann, M. E., and Jarzen, D. M. (1992). Community associations and structure in the Late Cretaceous vegetation of southeast Australasia and Antarctica. *Palaeogeogr. Palaeoclim. Palaeoecol.* **94**, 283–309.

Specht, R. L., Specht, A., Whelan, M. B., and Hegarty, E. E. (1995). *Conservation Atlas of Plant Communities in Australia.* Southern Cross University, Lismore, New South Wales.

BACTERIAL BIODIVERSITY

Erko Stackebrandt
DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH

GLOSSARY

domains The highest taxonomic rank defined to classify organisms into Archaea, Bacteria, and Eucarya, which differ from each other in fundamental genomic and phenetic properties.

endosymbiont Specialized form of symbiosis in which one partner thrives within cells, lumen, or tissues of the host organism; most obligate endosymbionts belong to the group of uncultured organisms.

epigenetic level Study of molecules which are the product of gene expression.

homology Denoting common ancestry. Structures, processes, sequences, behaviors, etc. are said to be homologous if there is evidence that they are derivations from a common ancestral structure. In molecular biology the term indicates a significant degree of similarity between DNA or proteins.

phylogeny Natural relatedness among life forms; the science of ordering the genealogy of organisms into a family tree.

prokaryotes Life forms that are members of the domains Archaea and Bacteria as opposed to Eucarya, comprising organisms with a cell nucleus.

rDNA Genes coding for rRNA that play a fundamental role in the translation process; the most thoroughly studied molecule in prokaryotic cells; used in comparative phylogenetic studies.

uncultured prokaryotes Organisms, the presence of which has been detected by molecular methods in the environment but they have not been cultured under artificial laboratory conditions.

BACTERIAL DIVERSITY comprises the total variability of prokaryotic life on Earth, covering all genomic, phenetic, phylogenetic, and ecological variations from the level of an individual strain to the community that evolved over a time span of probably more than 3.5 billion years. Determination of the extent of diversity is mainly restriced to cultured prokaryotes because the vast majority of strains are not yet accessible for subsequent research. Strains involved in obligate symbiotic and parasitic relationships with their eukaryotic hosts and free-living strains for which appropriate culture conditions have not been developed can at best be detected by molecular techniques. This article summarizes results of recent approaches that have broadened the bacteriologist's view about the immense richness of prokaryotic diversity.

I. RECOGNIZING GAPS IN KNOWLEDGE

Due to progress in methodologies and concepts, facets of biodiversity have no been covered equally in different areas in bacteriology, but progress must be viewed as a logical consequence of available technologies. Introduction of groundbreaking methods is usually followed by a period of increased knowledge in those fields for which the methods were developed. Numerous examples exist in microbiology, such as the elucidation of (i) the ultrastructure of the prokaryotic cell, following the development of the electron microscope; (ii) metabolic and biochemical pathways, following the introduction of the isotope label technique and enzymology; (iii) anaerobic organisms, following the development of appropriate anaerobic cultivation technology; and (iv) cell constituents, following the introduction of the amino acid analyzer, gas chromatography, high-performance liquid chromatography, and thin layer chromatography. Recently, the introduction of high-resolution gel electrophoresis led to one- and two-dimensional fingerprinting methods for proteins, ribonucleic acids (RNAs), and deoxyribonucleic acids (DNA); application of restriction enzymes, cloning strategies, and polymerase chain reaction (PCR) technology led to improved sequence analysis of genes and genomes; and cloning of PCR fragments of environmental DNA, *in situ* hybridization, and the development of gradient gel electrophoresis revealed a larger spectrum of prokaryotic diversity than previously known. The latter insights have been used to develop strategies in which DNA, isolated from environmental samples, is expressed to yield a spectrum of novel enzymes not detected in strains available from biological resource centers. As a result, taxonomists learned that the vast majority of prokaryotic species are still undetected, and physiologists, biochemists, and geneticists can expect to find novel lines of descent containing organisms that express fundamental deviations from currently known biochemical pathways. New models will help to better understand the structure and function of living matter and the role of prokaryotes in maintaining the biosphere.

The number of validly described species of animals, plants, and lower eukaryotes is approximately 400 times larger than the number of prokaryotic species (1,600,000:4,000). This fraction of bacterial and archaeal species is surprisingly low considering that prokaryotic species evolved eons ago, exploring and occupying any niche that has been investigated for the presence of prokaryotic organisms. Insects, on the other hand, which comprise more than 1 million species, evolved late in evolution—less than 600 million years ago during the Cambium. Certainly, differences in the species definition that exist between the biological species of higher evolved eukaryotes and the pragmatically defined prokaryotic species contribute to the tremendous discrepancy, but it has been shown convincingly that the vast majority of prokaryotic strains which are part of the free-living microbial community species have not been cultured. Endosymbiotic prokaryotes which are not free living but firmly associated with eukaryotic cells are another source of uncultured organisms. The limitation of recognizing the richness of microbial organisms is most likely due to the use of a restricted spectrum of enrichment media which selects for a very narrow spectrum of organisms that compete best under artificial laboratory conditions.

The discussion of diversity issues and the increasing awareness of the importance of microorganisms in maintaining the biosphere is embedded in discussion of the implementation of the articles of the Convention on Biological Diversity (CBD). The driving force behind the CBD has been the recognition of the possibility of a significant reduction or loss of biological diversity at source by human activities, and the preamble admits a general lack of information and knowledge regarding biological diversity. Nowhere is the lack of knowledge more acute than for microbial diversity (prokaryotes, fungi, yeasts, and heterotrophic and autotrophic protists) and there is a widespread appreciation among microbiologists that cultured microorganisms represent a very small, not necessarily ecologically important, fraction of natural microbial diversity.

Considering the total number of nucleotides per genome, far more recognizable diversity can be seen at the molecular level than at the epigenetic level. Compared to the total number of described species, the number of organisms analyzed in genome sequencing projects is still small (about 2.5%), but this portion will increase rapidly and genomic screening and data mining will develop into dominating biological disciplines in the future. However, taxonomists will not lose interest in the more traditional properties of a cell because of the demanding process of describing species and genera. The polyphasic approach to classification requires information about morphology, ultrastructure, metabolism, chemical structure of cell constituents, and genomic features. Thus, from more than 100 years of bacterial classification, an enormous wealth of phenotypic and epigenetic data have accumulated and are of benefit to bacterial taxonomy, which many scientists regard as the mother of biological sciences. However, taxonomists are concerned that this spectrum of diversity will have to be sacrificed for an approach that considers molecular/genomic data more appropriate for delineating taxonomic ranks. Indeed, this cut will happen sooner or later, but it does not mean that the avalanche of genomic data will bury all activities directed toward the elucidation of the phenotype.

II. UNRAVELING PHYLOGENETIC DIVERSITY

Whereas the past century unraveled the diversity of epigenetic properties, i.e., those characters that are the result of gene expression, the next century will prove a tremendous wealth of information on genomic properties. The terms genomics and reverse genetics have been coined for the strategy of obtaining value-added information from sequences of genomes and genes. Data mining and meganetworking programs will replace the simple search for similarities between a few homologous genes, and it can be expected that the importance of phenotypic traits will be reintroduced as a consequence of the desire to understand the horizontal and vertical flow of genes and their regulation and expression.

The combination of information of a (small) portion of the genome with phenotypic properties is nowhere exemplified better than in modern classification strategy of prokaryotes. The first component, the backbone of the system, is provided by the primary structure of homologous molecules which have accompanied the organisms since early evolution and they can be ranked according to their evolutionary history. The number and nature of sequence differences among proteins and genes coding for rRNA and proteins allow the recognition of pairs or groups of organisms which evolved from a common ancestor, and the order in which lineages have emerged in time facilitates decisions about the grouping of organisms.

The main problem that has emerged in phylogenetic studies during the past 10 years is the question of whether to place emphasis on the genealogical relationships derived from comparative analyses of a single homologous gene, a gene cluster, or even large parts of the genome. It has been demonstrated convincingly that the historic fate of ribosomal (r)DNA genes does not necessarily represent the fate of other genes because the phylogenetic branching patterns of different genes may show significant deviations. Analysis of the presence and position of genes in completely sequenced prokaryotic and eukaryotic genomes and horizontal gene transfer among genomes with a high plasticity has played an enormous role in designing and shaping early evolved organisms, and the chimerical genomic structure of descendants of early evolving organisms has been proven beyond doubt. The 16S rDNA is the most widely analyzed molecule in phylogenetic studies because of its alternating change of degrees of sequence conservatism that allows the recognition of most distant relationships and moderate and close relationships. The availability of a database comprising more than 10,000 sequences of prokaryotic strains makes it easy to either unambiguously affiliate a new isolate to one of the 4000 species or to postulate the finding of a new taxon in case no highly similar match is obtained with the isolate and a recognized species. Analysis of rDNA is the gold standard for analyzing phylogenetic relatedness; Schleifer and Ludwig (1989) provided excellent evidence that the rDNA data are closely matched by results of comparative analyses of other conservative molecules responsible for central cell functions. Second, the phylogenetic branching patterns provide a scenario in which clusters of related organisms also share a high portion of epigenetic properties. The resolution power of the 16S rDNA is restricted, however, because neither very early nor very recent evolutionary events are well resolved. Branching patterns are influenced not only by the treeing algorithms applied for inferring phylogenetic relationships but also by the size of the database. A phylogenetic tree is a dynamic construct which will to some extend change its topology with any new sequence added. Consequently, it is difficult for the scientist to judge whether the phylogenetic tree generated is a true reflection of the evolutionary history of the analyzed molecule. If the tree is generated for the purpose of making conclusions on the fate of genomic and epigenetic characteristics of the organisms concerned, conclusions should be underpinned by the analysis of additional evolutionary marker.

Increasingly, genes with a resolution higher than 16S rDNA are sequenced for determining intrageneric and intraspecies relatedness (e.g., genes *hsp65* and *hsp70* coding for heat shock proteins and *gyraseB*). Traditionally, and still required in the polyphasic approach to classification, DNA–DNA hybridization studies are performed at the level of highly related species and at the interspecies level, but the laborious experimental burden and lack of a cumulative database exclude wide application.

III. DIVERSITY EXTENDS TO THE STRAIN LEVEL

In the past, selected physiological reactions and computer-assisted techniques were applied to elucidate intraspecific relationships. This goal can also be achieved by generating patterns from DNA, RNA, and proteins that represent a strain-specific fingerprint or a bar code-type profile. Spontaneous mutations occur at an average rate of about 10^{-7} per gene and generation in prokaryotes but vary to a great extent in genetic loci. The extent to which strains of a species differ from each other also depends on the relative evolutionary time during which strains evolved from each other. Because prokaryotic species are man-made constructs established to facilitate taxonomy, strains included in a species may vary significantly from each other in phylogenetic depth and hence in the extent to which macromolecules diverged. The complexity of the patterns obtained depends on the size, length, and degree of conservatism of the macromolecule of choice and the tools used to cleave, amplify, hybridize, and separate these markers. Separation of DNA fragments by pulsed field electrophoresis, probing of defined genes with labeled probes, visualization of bands via computerized, laser-analyzed densitometer scanning, and the use of fully automated, reproducible techniques have improved the resolution and the monitoring part of the analyzes significantly. Pattern identification is rapid, discriminating, and applicable to any species for which DNA, RNA, and whole cell proteins can be isolated. Providing the potential for discrimination, isolates and reference strains that exhibit a high degree of pattern similarity can be considered related. These techniques complement traditional typing methods used mainly in the clinical environment, such as serotyping, biotyping, and phage typing. The decision of whether a strain with a unique pattern actually belongs to a described species or should be described as a new species requires more quantitative methods at the genomic level that allow one to measure the degree of relatedness.

IV. PHYLOGENETIC DIVERSITY AND DISTRIBUTION OF PHENOTYPIC TRAITS

The phylogenetic and phenotypic separation of the domain Archaea from the domains Bacteria and Eucarya (Fig. 1) is the most exciting result since the introduction of comparative sequence analysis by Woese and Fox in the mid-1970s. The presence of two prokaryotic domains in which members are defined by clearly different genomic and phenotypic properties has changed fundamentally the hypothesis on the dichotomy of life forms and revolutionized ideas about the evolution of the eukaryotic cell. Domains were introduced by Woese, Kandler, and Wheelis to denote that these primary lines of descent constitute higher entities than the traditional eukaryotic kingdoms. The most significant epigenetic differences among members of Archaea and Bacteria are the compositions of the cell wall and fatty acid and the modification pattern of tRNA.

Figure 1 schematically depicts the tripartition of the domain Archaea that guided the description of three kingdoms, the Euryarchaeota, the Crenarchaeota, and the Korarchaeota, for some uncultured organisms. No kingdoms have been described for the rich phylogenetic structure of the domain Bacteria, outlined in Fig. 2, because of the significantly large number of organisms and lineages involved, which are not always well separated from each other. The order at which these lineages evolved is of low statistical significance but the phylogenetic composition of organisms within these lineages can be recovered by analyses of other genes, such as those coding for 23 rRNA, 5S rRNA, ribosomal proteins, ATPase, elongation factors, and heat shock protein HSP70. In some lineages certain characteristics are indeed of phylogenetic significance, such as morphology and/or ultrastructural features (*Thermotogales, Planctomycetales, Verrucomicrobiales, Spirochaetales, and Myxobacteriales*), chemotaxonomic properties such as cell wall composition (*Thermotogales/Deinococcus*, clostridia, and *Actinobacteria*) or lack thereof (*Planctomycetales*), and physiology, i.e., the composition of the photosynthetic apparatus (*Chloroflexus, Chlorobiales*, and cyanobacteria). Most lineages, however, have a wide variation of morphological, chemical, ultrastructural, and biochemical diversity, some traits of which may have been acquired in the course of their evolution by horizontal gene transfer, whereas others may have evolved as a response to occupying new environmental niches (e.g., autotrophic and chemolitothrophic forms). Traits formerly believed to be of monophyletic origin—

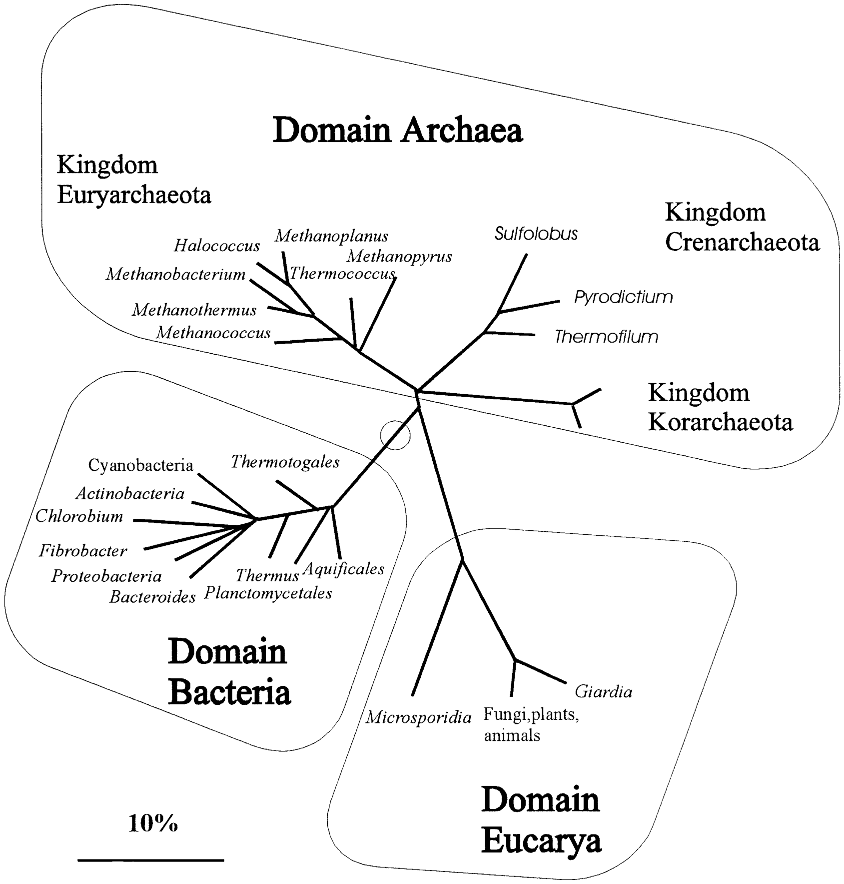


FIGURE 1 Schematic illustration of the main lines of 16S rDNA sequences showing the tripartition of the domains Bacteria, Archaea, and Eucarya. The circle indicates the approximate position of the root of the tree, making the domains Archaea and Eucarya phylogenetic neighbors. The scale bar indicates 10 estimated changes per nucleotide position.

and hence of taxonomic value—lost their significance in classification when their polyphyletic origin was demonstrated or when these properties were found to be of little genomic stability. In general, this is true for morphology, spore formation, the relationship to oxygen, the presence of a photosynthetic apparatus, gliding motility, and many other characters. In modern classification, these properties are no longer used as the sole basis for the description of higher taxa, such as families, orders, and classes; today, the rational for a higher taxon is primarily the distinct phylogenetic grouping of its members, whereas their phenotypic description may be rather broad.

A good example of the inability of phenotypic properties to serve as phylogenetic markers is provided by members of the class *Proteobacteria*, which includes the majority of Gram-negative bacteria. This class is highly diverse with respect to physiological and morphological properties. Morphological diversity ranges from simple spherical forms to the highly complex fruiting bodies of myxobacteria. Physiologies include chemolithoautotrophy, photosynthesis, fermentation, anaerobic respiration, and nitrogen fixation. To reliably affiliate a new isolate to a described genus by these properties is improbable; the chance of doing so is increased by the presence of unique biochemical properties, such as nitrate or ammonium oxidation (*Nitrobacter* in the α subclass and *Nitrosomonas* and relatives predominantly in the β subclass, respectively) and sulfate reduction (*Desulfovibrio* and relatives; δ subclass). Sulfur and sulfate oxidation as well as anaerobic photosynthesis are poor phylogenetic markers because members of *Thiobacillus* are found in the α, β, and γ subclasses and the sulfurless photosynthetic organisms occupy different sublines of descent within the α and β subclasses.

As a consequence of the recognized discrepancies

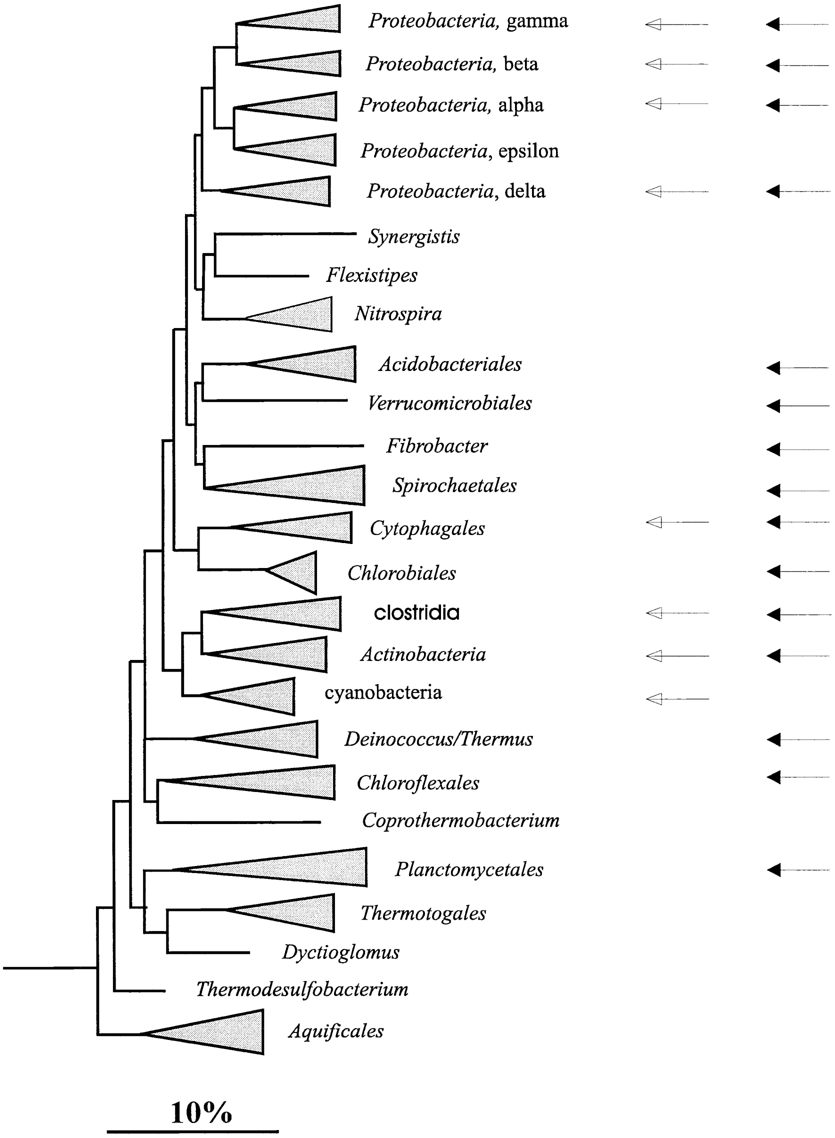


FIGURE 2 Relative order of the main lines of descent within the domain Bacteria based on 16S rDNA sequences. The horizontal portion of triangles indicates the phylogenetic depth of that lineage, whereas the size of triangles is not indicative of species numbers. Lineages from which symbionts evolved are indicated by open arrowheads. Lineages which encompass sequences of as yet uncultured free-living bacteria are indicated by solid arrowheads. The scale bar indicates 10 estimated changes per nucleotide position.

between the phylogeny of prokaryotic taxa and their previous taxonomic treatment, the recent phylogeny-oriented classification system of genera and higher taxa was developed to primarily match the 16S rDNA data. Newer textbooks have adopted the modern approach, in which, by providing the phylogenetic framework, the origin and evolution of certain phenotypic traits may be better understood than expressed by traditional superficial lumping.

V. THE DIVERSITY OF SYMBIOTIC PROKARYOTES

Endosymbiotic associations, recognized more than a century ago, initially concentrated mainly on the elucidation of the origin of chloroplasts. Later, the importance of eukaryote–prokaryote relationships was recognized for associations between plants and nitrogen-fixing bacteria, e.g., legumes and *Rhizobium* species, monocots and *Azospirillum* species, *Casuarina* sp. and *Frankia* sp., ferns and *Anabena* sp., and the prokaryotic origin of plant mitochondria. However, most endosymbiotic relationships between microorganisms and their eukaryotic hosts were mainly descriptive and the taxonomic affiliation of the vast majority of uncultured microorganisms remained virtually unknown. In contrast, microbial partners participating in nonobligate symbiotic relationships were identified long before the molecular era; there are numerous associations for the ectosymbiotic microbiota of the rumen, intestine, gut, and skin.

A. Identification of Symbionts

Most endosymbiotic bacteria are defined as "as yet uncultured" organisms, many of which do not exist as a pure culture within the host's tissue. Hence, genes coding for rRNA cannot be isolated selectively but must be identified within clone libraries consisting of PCR-amplified rDNA using prokaryote-specific PCR primers and total DNA extracted from the plant, animal, or even the prokaryotic cell. If the host contains a single symbiotic partner only, the clone library will consist exclusively of the one unique rDNA insert; if the association is more complex, the clone library will contain phylogenetically different inserts. Authentification of the symbiont and verification of the location of the putative endosymbiont within the host's tissue is performed by fluorescence *in situ* hybridization techniques as elegantly developed by Stahl, Amann, and coworkers. In most cases the 16S rDNA sequence will provide sufficient unique nucleotide stretches to allow generation of symbiont-specific oligonucleotide probes.

Not until 1982 was *Prochloron didemni*, the endosymbiont of the ascidian *Lissoclinum patellum*, identified by molecular techniques. Ten years later, a wealth of information was available on the molecular phylogeny of symbionts and endosymbionts from a broad spectrum of eukaryotic hosts, ranging from protozoa to vertebrates. The availability of a large database, consisting of thousands of 16S rDNA sequences of free-living prokaryotic species, facilitates the search for the phylogenetic affiliation of the more than 500 16S rDNA sequences available for host-associated bacteria. Fundamental questions about the identity of the prokaryotic partner, the evolution of symbiotic relationships, and the mechanisms of symbiont transmission can now be addressed.

B. The Evolutionary Origin of Symbionts

Most endosymbionts investigated to date originate from ancestors within the domain Bacteria, in which they are found in a few main lines of descent (open arrowheads in Fig. 2). The ability to thrive in certain anaerobic protozoa of the genera Metopus, Plagiopyla, and Trimyema appears to be widespread among methanogenic Archaea (kingdom Euryarchaeota). No endosymbionts have been described to share a common ancestry with those prokaryotes which define the most deeply branching lineages, such as the Crenarchaeota, domain Archaea, and *Aquificales, Thermotogales*, and other branches comprising thermophilic and phototrophic organisms of the domain Bacteria. Because symbionts and nonsymbionts share more than 80% 16S rDNA sequence similarity it can be concluded that the invasion of the eukaryotic host by prokaryotic cells must have occurred less than 2 Gy ago. Figure 3 shows recent evolutionary events by plotting geological time against 16S rDNA similarity values determined for the origin of organisms defined by key physiological types (i.e., oxygen-generating photosynthesis by cyanobacteria, fermentative metabolism in facultative anaerobic bacteria, and origin of respiration chain in aerobic bacteria). Thus, as derived from 16S rDNA similarity values, the origin of endosymbionts correlates with the origin of the eukaryotic cell and endosymbiosis has occurred repeatedly in (perhaps all) eukaryotic lineage.

In order to study the history of symbiotic associations, the phylogenetic trees of hosts and their symbionts should be compared. Few data are available, with the most convincing study being that on the endosymbi-

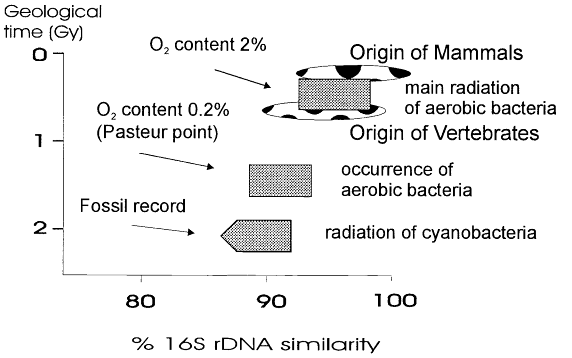


FIGURE 3 Correlation plot of geological time (as deduced from the fossil record and oxygen content) and 16S rDNA similarity values. Only the past 2 billion years are shown. The origin of aerobic bacteria and symbiotic bacteria correlates with the increased content of oxygen in the atmosphere and the eukaryotic cell, respectively (modified from Stackebrandt 1995).

onts of aphids. The topology of the symbiont *Buchnera aphidicola* tree is completely concordant with host phylogeny based on morphology. The fossil and biogeographic time points for the aphid phylogeny have been used by Wilson and Bauman to calibrate the 16S rDNA of the closely related endosymbionts (>8% similarity). The value of 1% fixed substitutions per 25–50 million years determined for the symbionts of aphids is similar to the value of 1% per 50 million years determined on the basis of a broader range of nonobligate symbiotic relationships (e.g., Rhizobium/legumes, Photobacterium/fish, and enterobacteria/mammals) and to the value of 1% per 60 million years for the past 500 million years (Fig. 3).

The host's advantage of the association has been unraveled in a few cases, such as the removal of hydrogen produced from hydrogenosomes of ciliates by archaeal methanogenic endosymbionts, provision of nutritional carbon to the host bivalves by sulfur-oxidizing gill symbionts, or essential amino acids to aphids by their endosymbionts. Application of the PCR techniques has allowed the elucidation of the transmission route of symbionts in ovaries, testis, and gill tissue of tropical lunicid bivalves and deep sea bivalves.

C. Phylogenetic Affiliation of Endosymbionts

The majority of endosymbionts cluster phylogenetically most closely with Gram-negative free-living bacteria. It can therefore be deduced that Gram-negative bacteria are the most successful candidates for forming symbiosis, including obligate endosymbiotic associations. A few nonproteobacterial symbioses have been described, such as those found between wood-eating cockroaches and termites and spirochetes and Gram-positive host-associated bacteria, such as the microparasite *Pasteuria penetrans* (in root-knot nematodes), *Epulopiscium* (a fish symbiont), and *Frankia* (nitrogen fixing on *Casuarina* and relatives). The association between *Sphingobacterium comitans*, a member of the *Bacteroides/Flavobacterium* phylum, and the myxobacterium *Chondromyces crocatus* is an example of a prokaryote–prokaryote symbiosis. Endosymbionts of Archaea are members of the kingdom Euryarchaeota and they are, in contrast to those of Bacteria, organisms containing a Gram-positive cell wall. These anaerobic symbionts, originating from ancestors of the families *Methanosarcinaceae, Methanomicrobiaceae*, and *Methanocorpusculaceae*, have been identified in several genera of termites and protozoa. The basis of the symbiotic interaction appears to be hydrogen transfer from host to endosymbiont, which can use the gas for methanogenesis.

1. Symbionts of the Proteobacteria

Proteobacteria embrace organisms known to have close associations with eukaryotic hosts as pathogens, parthenogenesis bacteria, incompatibility bacteria, symbionts, and organelles (such as the mitochondria of plants which evolved from α proteobacterial ancestors). Endosymbionts and nonobligatory associates of the same host may belong to different phylogenetic groups, indicating that the same host is susceptible to more than a single invasion process and that not all symbiotic relationships result in obligate endosymbiosis.

a. Symbionts of the α Proteobacteria

The α subclass of Proteobacteria contains a wide spectrum of organisms that are closely associated with eukaryotic cells. Prime examples are members of the nitrogen-fixing genera *Rhizobium, Sinorhizobium, Bradyrhizobium, Mesorhizobium*, and *Azorhizobium*. Pathogens include *Afipia, Brucella, Bartonella, Rickettsia, Ehrlichia, Orienta*, and *Anaplasma*. A highly related cluster of host-associated organisms, including symbionts of insects, cytoplasmatically inherited bacteria such as the parthenogenesis bacteria (PB), and the cytoplasmatic incompatibility bacteria (CIB) is related to *Wolbachia pipientis*. Other members of this group are the PB and CIB of *Culex* and *Drosophila*. Because of the degree of 16S rDNA similarity among cultured strains, these symbionts must be considered members of the same species. These relationships demonstrate that the common ancestor of *Wolbachia* and relatives invaded a broad spectrum of insect hosts which, as shown by the high values of up to 99% 16S rDNA similarity, must have occurred recently in evolution.

b. Symbionts of the β Proteobacteria

Members of the β subclass of Proteobacteria encompass a wide range of mainly pathogenic plant- and animal-associated bacteria, such as members of *Burkholderia, Azoarcus, Ralstonia, Bordetella, Kingella, Eikenella, and Neisseria*. Also included in this subclass is the kinetoplast of *Crithidia*. The diversity of endosymbionts, however, is rare and restricted to the endosymbionts of the mealy bugs, which are moderately related members of the genera *Ralstonia* and *Burkholderia*.

c. Symbionts of the γ Proteobacteria

By far the higher number of endosymbionts of insects and vertebrates are members of the γ subclass of Proteobacteria. In addition, this taxon contains a wide spectrum of animal and human pathogens and nonobligate symbionts such as members of *Enterobacteriaceae, Legionellaceae, Pasteurellaceae, Vibrionaceae, Pseudomonas* (sensu stricto), and *Acinetobacter*. Many endosymbionts cluster according to the phylogenetic rank of their hosts, which may be indicative of coevolution events: (i) The primary endosymbionts of giant ants, aphids, tse-tse, and the sweet potato white fly form a phylogenetically coherent cluster within the radiation of enterobacteria; (ii) the symbionts of fish with light organs, such as the deep-sea anglerfish and the flashlight fish, cluster with different members of *Vibrio*; (iii) two moderately related subcluster consist of the gill symbionts of bivalves. These organisms are remotely related to methylotrophic bacteria. The latter symbionts are sulfuroxidizing organisms which provide their hosts with nutritional carbon. Nitrogen-fixing symbiosis are not restricted to plants but the gland of an invertebrate shipworm contains large numbers of γ proteobacteria which possess the ability to digest cellulose and fix nitrogen.

VI. THE DIVERSITY OF THE UNCULTURED FREE-LIVING ORGANISMS

For more than a century, assessment of prokaryotic species has been evaluated by the culturing approach. The number of different growth media is unknown, but all aimed to recover the largest possible diversity of organisms. However, it is not the enrichment and isolation but rather the lack of cheap and reliable molecular identification methods that still slow microbiologists in their attempts to classify strains to the species level and to describe new prokaryotic species. The number of novel strains that has been eliminated during the isolation process, or which were only included in a biased search for specific properties before they disappeared in nonpublic resource collections, cannot be counted. It is unknown how many strains have been investigated in parallel. Although this problem may one day be overcome by more facilitated species definitions and the availability of a global network of biological information, microbiologists are currently confronted with the problem that there is a remarkable difference in the number and morphology of organisms in natural samples with enrichment cultures and isolated colonies. Staley and Konopka introduced the phrase “great plate anomaly” to indicate that only a small fraction of prokaryotic species observed under the microscope will grow under artificial laboratory conditions.

A. The Vast Majority of Prokaryotic Species Have Not Been Cultured

Pace and colleagues first suggested that rRNA sequences could be used to characterize natural communities without the need to culture. The delay in publishing the first studies by the research groups of Giovannoni and Ward was largely due to the need to develop robust and simple technologies whereby 16S rRNA sequences could be recovered from complex mixtures of environmental nucleic acids and then individually sequenced. Since then, there have been numerous 16S rRNA- and 16S rDNA-based studies in which sequences were analyzed to explore microbial diversity in different environments. Most of these studies differ in details of methodologies, such as isolation of nucleic acids, PCR amplification conditions, and source of cloning vectors and ligation enzymes, but in one aspect the outcome of all of these studies was similar: The vast majority of the more than 1200 environmental partial 16S rDNA sequences, deposited in public databases, were not identical with and often not even similar to the homologous sequences of described species accessible in the extensive databases of cultured bacteria (Fig. 2, solid arrowheads). Also, the sequences were rarely identical to sequences obtained from strains that were isolated in parallel to the molecular work from the same environment (Fig. 4). As judged from the low degree of sequence similarity, one could even conclude that many sequences are indicative of the presence of higher taxa. This finding reinforced the previously mentioned idea that the vast majority of species have not yet been isolated.

Additional molecular techniques have been developed for understanding the composition of microbial communities that reach beyond the mere assessment of phylogenetic relationships of clones and strains. Among others, methods comprise (i) the application of gradient electrophoresis of PCR-amplified 16S rDNA sequences to facilitate recognition of changes of populations in time and space, (ii) the development of biological probes to detect the presence of genes encoding metabolic enzymes or to identify bacterial species directly in an environmental sample, (iii) the development of biosensors to determined microprofiles of inorganic compounds, (iv) flow cytometry and cell sorting to enumerate and separate groups of organisms according to size and taxon specificity, (v) subtractive hybridization to facilitate comparative analysis of environmental samples, and (vi) extension of the database to include genes other than 16S rDNA (e.g., *nif* genes).

The environments discussed in the following sections were selected because they provide the largest database of phylogenetic information on uncultured organisms. Many other environments have been investigated, but the overall picture of prokaryotic diversity is less focused; these include, fresh water, paddy fields, marsh oil, marine plants, bioremediation sites, bioleach reactors, or the multistructured associations between prokaryotes and eukaryotic cells.

1. The Marine Environment

a. The Uncultured Archaea

The use of archaeal-specific 16S rDNA PCR primers and subsequent analysis of clone libraries from DNA of oxygenated coastal surface waters and oligotrophic open-ocean samples by the groups of DeLong and Fuhrman revealed the widespread occurrence of two types of archaeal diversity. Analyses of different marine sites, including surface and subsurface waters of the Pacific, Atlantic, Antarctic coastal waters, offshore slope regions, and a deep-sea marine holothurian, have subsequently confirmed the affiliation of archaeal sequences to these two groups. Clone sequences of the first group contained 16S rDNA genes of Crenarchaeota organisms found at depths >100 to 4800 m, whereas others were found at different marine sites at more shallow depths. Organisms from which sequences originated constitute a significant component of approximately 5–14% of the marine picoplancton assemblages and were novel and closely related to the archaean "Crenarchaeum symbiosum," a symbiont of the sponge *Axinella mexicana*. At a lower level of relatedness these sequences form a separate branch at the basis of the kingdom *Crenarchaeota*, where they show some distant relatedness to a group of sequences isolated from sediment organisms of the Obsidian Pool, Yellowstone.

A few sequences have been described to belong to the second sequence group originating from Euryarchaeota and are distantly related to the terrestrial species *Thermoplasma acidophilum*. Except for a holothurian archaeal sequence, they have been retrieved from the same environment from which the crenarchaeotal sequences were obtained, but mostly at lower depth (0–100 m). The high sequence similarity of members of this group from geographically separate sites contrasts the finding by Munson and colleagues, who demonstrated the high degree of phylogenetic diversity between archaeal sequences retrieved from 16S rDNA clone libraries generated from material sampled in marsh sediment samples and adjacent vegetative marshland in the United Kingdom. In this study, clone sequences formed about 15 different phylogenetic groups, each of which was highly to distantly related to cultivated Euryarchaeota species.

The origin of marine archaea is unclear because no representative has been cultured. One may argue that these prokaryotes originate from dormant stages of released commensales or symbionts of marine invertebrates. This view, however, is contradictory to the high cell number of living organisms of this group—up to 14% of the total community. Disturbed deep-sea sediments, which have not been investigated, may be another possible source for these organisms. In contrast, euryarchaeal clone sequences retrieved from surface water material may be coastal and even terrestrial origin, considering that there is a specific relationship of these sequences to some of the DNA retrieved from coastal salt marsh.

b. The Uncultured Bacteria

Phylogenetic analysis of rDNA from phytoplankton of the Sargasso Sea by Giovannoni, Britschgi, and coworkers revealed that the majority of the mostly eubacterial sequences were novel to systematists. Some clones represented oxygenic phototrophs and could be assigned to the *Synechococcus* group of the *Cyanobacteriales*. Groups of related sequences of mainly the α and γ Proteobacteria were identified and defined as SAR (Sargasso) groups, showing distant relatedness to cultured bacteria of *Shewanella, Vibrio*, and *Oceanospirillum*. Analysis of the phytoplankton from the north Pacific near Hawaii by Schmidt and colleagues led to the unexpected result that the population was very similar to that of the Atlantic Ocean. Some sequences were similar to those of common cultured marine organisms (e.g., *Vibrio, Pseudoalteromonas*, and *Chromatium*), whereas others represented novel lineages which were distantly related to the *Fibrobacter* group and *Chlorobium*.

Following studies extended the range of sampling sites in the Atlantic and Pacific Oceans, the Antarctic Sea, and the Mediterranean Sea. Basically, the findings of the first studies were confirmed in that many of the new sequences were highly related, although not identical, to those defined earlier, irrespective of the location. The ecological role of the hitherto undescribed organisms remains unresolved, although their general physiological capacity is probably similar to that of described species because α Proteobacteria encompass mainly lithotrophic and oligocarbophilic organisms, whereas γ Proteobacteria and cytophages exhibit strong hydrolyzing and degrading capacities. The few sequences of Gram-positive bacteria are probably of terrestrial origin because they are mainly found in coastal regions and in sediments which must be regarded as a deposit mainly for endospore-forming organisms.

B. Hot Spring Environments

Research in hot spring environments has been stimulated by the discovery of *Thermus aquaticus* as a species of high biotechnological value. The community of mats especially provided an excellent comparison of microbial composition as assessed by selective culture techniques introduced by Brock and Castenholz. Results of the sequence investigations by Ward, Weller, and colleagues clearly demonstrated that even such a rather closed system inhabits a phylogenetically diverse community. Consistent with the findings in the marine environment, none of the recovered sequences closely resembled sequences from cultured taxa isolated from similar environments. Sequences of cultured organisms believed to constitute a major component of the mat community, such as *Synechoccous lividus* or *Chloroflexus aurantiacus*, were not recovered. Analysis of prokaryotic DNA from hot springs located in Yellowstone National Park by Pace and coworkers revealed an unexpectedly large number of distinct bacterial sequences, indicative of novel main lines of descent. Archaeal sequences were determined to be specifically related to sequences from the Crenarchaeota. A few sequences showed high similarities with 16S rRNA sequences of cultivated Archaea, (e.g., *Desulfurococcus mobilis, Pyrobaculum islandicum*, and *Thermofilum pendens*) but were not identical to any. The archaeal sequences from the hot spring environment were not closely related to the novel archaeal sequences retrieved from the marine environment, which appear to possess a similar position intermediate to the Crenarchaeota. Results of these studies indicated that the domain Archaea possesses a third line of descent, the kingdom Korarchaeota, suggesting that not only the phylogenetic but also the physiological diversity of the Archaea are significantly larger than reflected by the few cultured representatives.

C. Soil Environment

In contrast to the early extensive work on different marine sites and hot springs, determination of microbial biodiversity in soil was delayed until appropriate methods were developed that circumvented methodological difficulties such as the isolation of PCR-able DNA from humic acid-containing soil and semiquantitative cell recovery. Despite these shortcomings studies by Liesack and Triplett clearly demonstrated that soil samples from the Southern and Northern Hemispheres contain a rich and varied bacterial flora, including nonthermophilic archaeal members from the kingdom Crenarchaeota.

The first soil sample investigated by sequencing and probing 16S rRNA genes was located in a subtropical, moderately acidophilic, and forested environment in Queensland, Australia. Unexpectedly, only a few sequences were obtained from commonly isolated soil organisms, such as *Streptomyces*, although members of this genus were cultured in large numbers from the same soil sample. The reasons for the low representation of Gram-positive bacteria in libraries generated with universal 16S rDNA primers are not known, but this may be explained by the cell wall structure of resting and dormant cells which fail to disintegrate under mild enzymatic lysis. By using actinomycete-specific primers, Embley and colleagues were indeed able to demonstrate the presence of a rich diversity of these taxa in soil. Alternatively, Gram-positive bacteria may be a minor (numerically) component of the soil flora, with selective isolation exaggerating their numbers. Some Australian soil sequences were closely related to those of nitrogen-fixing species of the α Proteobacteria, but the majority of clone sequences represented novel groups that were only remotely related to known taxa, e.g., *Planctomycetales, Actinobacteria, Verrucomicrobium, Acidiphilium*, and *Thiobacillus acidiphilius*.

Several actinobacterial clone sequences, retrieved from rDNA of different soil types and different geographical locations, formed two clusters grouping remotely with members of *Rubrobacter, Acidicrobium*, and *Atopobium*. In addition to their presence in Australian forested soil, their occurrence was verified in a hot spring (Australia), geothermally heated soil (New Zealand), paddy fields and soybean fields (Japan), cultivated soil (Mexico), peat bog and garden soil (Germany), grassland soil (The Netherlands), and forest soil (Finland) (Fig. 4). These sequences formed fractions of 1–23% of the respective clone libraries and one of these organisms constituted about 6% of the metabolically active part of a Dutch grassland soil community as shown by ribosomal RNA analysis. It can be deduced that these uncultured organisms

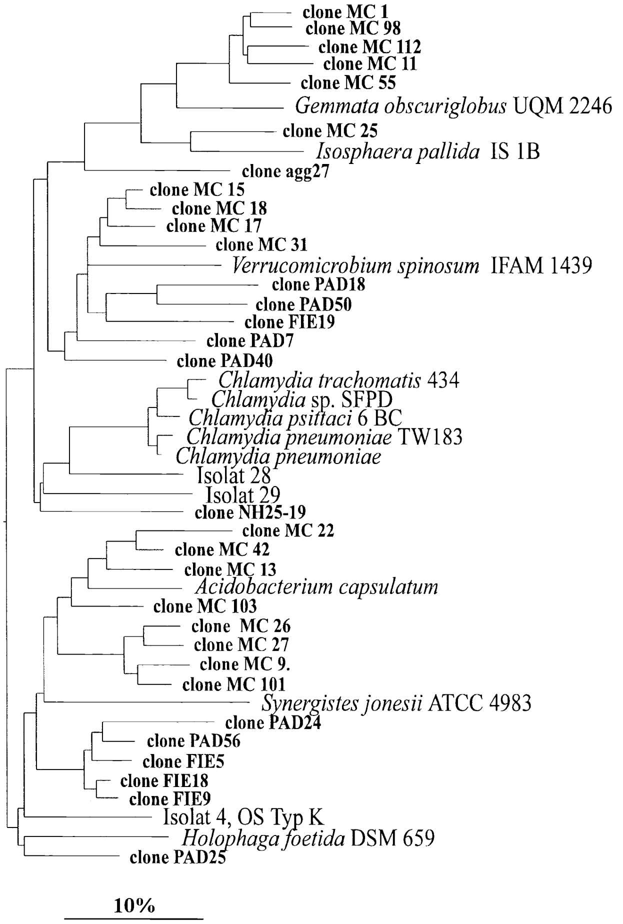


FIGURE 4 Distribution of 16S rDNA clone sequences, representing uncultured bacteria, within the radiation of 16S rDNA sequences of some cultured bacteria. Origin of clone sequences: MC, Australian soil; PAD, Japanese paddy field; FIE, Japanese soy bean field; NH, north Pacific Ocean. The scale bar indicates 10 estimated changes per nucleotide position

are distributed worldwide and play a physiologically important role in the soil ecology.

VII. CONCLUSIONS

Despite tremendous progress in the elucidation of prokaryotic diversity, many pitfalls have been identified which influence the composition of sequences in a clone library and hence these data can be used neither to quantify nor to qualify the composition of communities. Any estimation of the relation of cultured and as yet noncultured organisms is nothing more than a guess and not supported by scientific data. Most environmental analyses have revealed a heterogeneous mixture of deep and shallow branching lineages, very few of which have shown close relationships to cultured taxa. Many lineages are very closely related to each other, some of this diversity is due to microheterogeneity at the level of rrn operons within a single cell, whereas others may represent true strain diversity.

Although the contribution of rDNA and rRNA to microbial ecology must be considered significant, one molecule alone cannot nearly cover all facets of microbial ecology. Not only must the function of an ecosystem be deduced from analyses of genes expressed through rRNA, mRNA, and proteins, and not only should the network of broad physiological interactions be verified by *in vitro* reconstitution of isolates but also all these strategies must include data on thorough physical and chemical analysis of the natural sample. Because ecological interactions are performed by strains and not by species, the 16S rDNA in not the appropriate tool to unravel the diversity of this high phylogenetic level. Strain diversity may thus be one or more magnitudes higher than mirrored by the analysis of such an evolutionary conserved gene. Ecological niches must be defined and the difficulty in doing so is increased with the complexity of the sample. Soil samples, for example, are homogenized in that during the isolation of DNA many individual microniches with their individual populations are destroyed. In order to understand ecological interactions, population sizes must be known, strain richness and strain abundance must be assessed, the biochemical diversity of strains must be recognized by applying functional probes for the detection of the expression of specific genes, and physical probes must be applied for the assessment of the chemical and physical conditions of the environment. Modern environmental studies must maintain the isolation component because biotechnological exploitation of cosmid libraries is as desirable as the increase in knowledge of the evolution, phylogeny, and ecology of pure cultures.

See Also the Following Articles

ARCHAEA, ORIGIN OF • BACTERIAL GENETICS • EUKARYOTES, ORIGIN OF • MICROBIAL BIODIVERSITY, MEASUREMENT OF • THERMOPHILES, ORIGIN OF

Bibliography

Embley, T. M., and Finlay, B. J. (1994). The use of small subunit rRNA sequences to unravel the relationships between anaerobic ciliates and their methanogen endosymbionts. *Microbiology (Reading)* 140, 225–235.

Ochman, H., and Wilson, A. C. (1987). Evolution in bacteria: Evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74–86.

Pace, N. A. (1997). A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.

Schleifer, K. H., and Ludwig, W. (1989). *Phylogenetic relationships among bacteria.* In *The Hierarchy of Life* (B. Fernholm, K. Bremer, and H. Jörnwall, Eds.), pp. 103–117. Elsevier, Amsterdam.

Service, R. F. (1997). Microbiologists explore life's rich, hidden kingdoms. *Science* **275**, 1740–1742.

Von Wintzingerode, F., Göbel, U., and Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis. *FEMS Microbiol, Rev.* **21**, 213–229.

Ward, D. M., Ferris, M. J., Nold, S. C., and Bateson, M. M. (1998). A natural view of microbial diversity within hot spring cyanobacterial mat communities. *Microbiol. Mol. Biol. Rev.* **62**, 1353–1370.

Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**, 221–271.

Zuckerkandl, E., and Pauling, L. (1965). Molecules as documents of evolutionary history. *J. Theor. Biol.* **8**, 357–366.

BACTERIAL GENETICS

Michael Travisano
University of Houston

GLOSSARY

conjugation Direct cell-to-cell transfer of DNA.
operon The basic unit of bacterial gene organization, containing transcription start and stop sites.
plasmid DNA molecules that replicate independently of the major bacterial chromosome(s).
regulon A group of operons jointly regulated by one factor.
transduction Viral-mediated genetic recombination.
transformation Uptake and integration into the host genome of free DNA from the environment.

PRIOR TO PHYLOGENETIC RECONSTRUCTION using DNA sequence data, all prokaryotic organisms were called bacteria and species were determined primarily by phenotypic characteristics such as nutrient metabolism. With the advent of molecular systematics, prokaryotes have been divided into two lineages: Eubacteria and Archea. Eubacteria and Archea are evolutionarily distant and are two of the three domains of life (Eukarya is the third) (Fig. 1). The evolutionary divergence among the three groups is large and phylogenetic trees based on 16S and 18S ribosomal trees indicate that Eubacteria are more distantly related to Eukarya and Archea than the latter two are to each other. However, data from other genes give different outcomes and the precise evolutionary relationship of Eubacteria, Archea, and Eukarya is a topic of intense research activity. This article focuses exclusively on the genetics of Eubacteria.

I. TAXONOMY

Eubacteria (forthwith denoted simply bacteria) are a genetically highly diverse group of organisms and have been found in virtually every environment examined. Bacteria have been isolated from hot springs, deep-sea hydrothermal vents, inside Antarctic rocks, and as endosymbionts in many organisms. There is greater genetic diversity among Eubacteria than in beetles, insects, or arthropods—taxa notable for genetic diversity. This diversity results from both a long evolutionary history (more than 3.5 billion years) and the effects of both selection and genetic drift. Unlike eukaryotes, much of the genetic diversity of bacteria is largely unknown with the vast majority of species not characterized. Until the advent of molecular tools, only those bacteria culturable in the laboratory could be easily studied, and such bacteria may be no more than 0.1% of the total diversity.

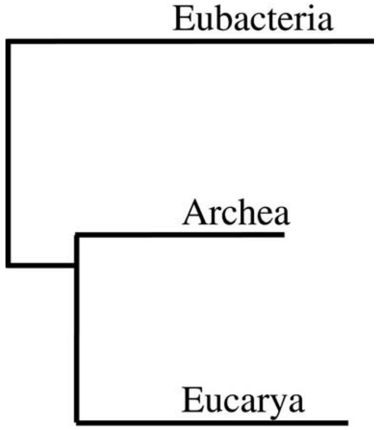


FIGURE 1 Divergence of the three major domains of life.

Gram staining has historically been used to classify bacteria into two broad groups. Bacteria which retain the crystal violet–iodine complex used in gram staining are gram positive and those which do not are gram negative. Differences in gram staining result from differences in cell wall structure, and as such gram staining generally does correlate well with actual phylogenetic differences. However, based on 16S rDNA trees, there are 19 currently recognized main bacterial lineages (Table I). Of the 19 lineages, three phyla (Proteobacteria, gram-positive bacteria, and Cyanobacteria) comprise more than 90% of the known genera and contain the majority of described species. Most characteristics which previously had been used in bacterial taxonomy (e.g., nitrogen fixation, prototrophy, sulfur reduction, motility, budding reproduction, and spore formation) are of little value in determining phylogenetic relationships. Similarly, phylogenetic distance is not a good predictor of the mode of life in bacteria and extraordinarily different bacterial phenotypes can be closely related.

Assessing bacterial diversity is difficult because the biological species concept cannot be easily applied to most bacteria. According to the biological species concept, a species is made up of those individuals which interbreed in nature. Excluding bacterial species undergoing natural transformation, recombination is too infrequent to be a defining species characteristic for bacteria. The commonly used alternative adopted by bacterial taxonomists is based on the tendency for double-stranded DNA to "melt" into single strands when heated. Homoduplex DNA, which is double-stranded DNA from the same bacterial strain, will tend to have a higher melting temperature than heteroduplex DNA, which is double-stranded DNA from different strains, due to DNA sequence differences between the two strains. The greater the sequence differences between strains, the lower the melting temperature of the heteroduplex DNA. Two bacterial strains are considered to be members of a single species if the difference between the melting temperatures of the double-stranded homoduplex DNA and heteroduplex DNA is not more than 5°C and if they share no less than 70% identical DNA base pairing (sequence similarity). This alternative species criteria reflects a sequence divergence of no less than 7 or 8% and would group together naturally transformable bacterial species that can be differentiated by the biological species concept. If applied, the criteria would also fail to differentiate many eukaryotic species, such as humans, chimpanzees, and gorillas. However, based on such criteria one would estimate that there are between 10^9 and 10^{12} bacterial species (Dykhuizen, 1998).

II. GENETIC ARCHITECTURE

Genetic architecture refers to the physical and informational structure of the genome. How is the genome organized and how is the encoded information expressed? Across all bacterial lineages much of the genetic basis of heredity is the same. Bacteria are haploid and all genetic material in bacteria is encoded by double-stranded DNA molecules using four bases: adenine, guanine, cytosine, and thymine. The sequence of bases in DNA provides the information necessary for RNA and protein structure. Nevertheless, genetic differences among bacteria abound and it is these differences that allow bacteria to be reliably differentiated. In contrast, although phenotypic differences are also prevalent, they are of substantially less value in differentiating bacteria, and over evolutionary timescales bacterial phenotypes are highly variable. The ease at which evolutionary transitions can be made and the great variety of habitats in which bacteria are found indicate that the bacterial genetic architecture, as a whole, is able to respond to selection relatively easily. The persistent and sometimes large differences in genetic architecture among bacterial species appears to engender minimal constraints on bacterial diversification.

Against this backdrop of diversity, *Escherichia coli* is the best characterized bacterium. Much of the initial work on bacterial metabolism and genetics was done with *E. coli*, and all subsequent work has used *E. coli* as a comparison. The article will follow a similar convention.

A. Physical Structure

The physical structure of bacterial genomes is strikingly divergent from that of eukaryotes. Even so, the diversity among bacterial species in genome size, organization, and method of replication may be as large or larger.

1. Size

Bacterial genomes are typically small: *E. coli* strain K12 has a genome size of 4.6 million bases (mb) in comparison with the ~12 mb of the eukaryotic yeast *Saccharomyces cerevisiae* and 3300 mb for humans (Fig. 2). The small size of bacterial genomes reflects the simplicity of the bacterial cells but also appears to be due to the lack of "junk" DNA in bacterial genomes. Bacterial genes are generally packed closer together than in most eukaryotes with relatively few intergenic spacer regions. In *E. coli*, ~85% of the chromosome is composed of RNA and protein coding regions. Among bacteria, smaller genome size is often found in parasites and obligate endosymbionts. For example, *Mycoplasma genitalium*, which is found in the human genitourinary tract, has the smallest genome capable of independent replication at ~0.6 mb, whereas the free-living soil bacterium *Myxococcus xanthus* has a genome of approximately 9 mb. Endosymbionts and parasites can lose some metabolic pathways and depend on their host for those gene products. Chlamydia, which are obligate animal parasites, depend on their host for energy generation because they are incapable of either synthesizing adenosine triphosphate or reoxidation of reduced nicotinamide adenosine dinucleotide phosphate. However, many pathogenic bacterial species do not have especially small bacterial genome sizes. *Mycobacterium tuberculosis*, the bacterium causing tuberculosis, has a genome size of 4.4 mb, which is slightly larger than the 4.2-mb genome of free-living soil bacterium *Bacillus subtilis*.

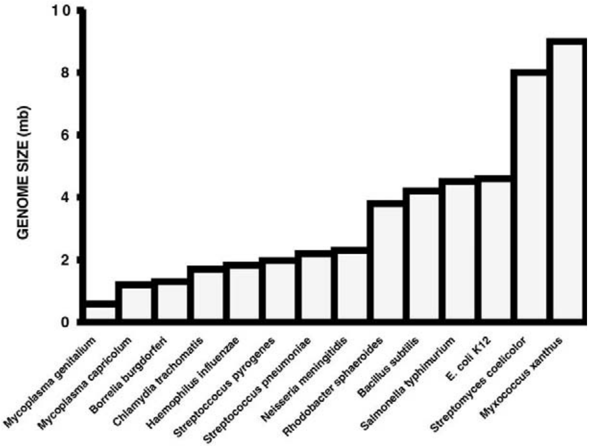


FIGURE 2 Genome sizes of some representative bacteria.

2. Chromosomes

Bacterial chromosomes are located in a nucleoid, a distinct cytoplasmic structure, in which double-stranded DNA is coated with histone-like proteins. Most bacteria appear to have a single large circular chromosome, but this is not universal. Many species have multiple chromosomes, such as *Rhodobacter sphaeroides*, which has one chromosome of 3.0 mb and a second of 0.9 mb, and *Burkholderia cepacia*, which has three chromosomes with respective sizes of 3.6, 3.2, and 1.1 mb. Other species have linear chromosomes, such as the spirochete *Borrelia burgdorferi*, which is the causative agent of Lyme disease, and the gram-positive *Streptomyces coelicolor*. In addition, many bacteria contain extrachromosomal elements such as plasmids (see Section II,A,3). The genome of *B. burgdorferi* is composed of an ~0.9-mb linear chromosome and ~19 linear and circular plasmids with a total size greater than 0.56 mb.

The type of chromosome topology has strong effects on chromosome replication. Unlike in eukaryotes, bacterial chromosomes have a single site for initiation of replication. In *E. coli*, replication of the single circular chromosome is initiated at the *oriC* site (*ori*gin of replication) and proceeds in both directions in the same semiconservative fashion as in eukaryotes. Gene organization of the origin of replication region is evolutionarily conserved across many bacteria lineages (e.g., *E. coli* of the phylum Proteobacteria and gram-positive *B. subtilis*) suggesting early fixation of a general mechanism for chromosome replication. DNA replication continues around the circular chromosome until both replication forks meet in the terminus region, which forms a barrier to replication fork progression. Specific AT-rich sites (Ter sites) within the terminus region impede replication from one direction, but not the other, and are oriented so that genome replication can be completed (Fig. 3). In contrast, the linear chromosome of *B. burgdorferi* is bounded by telomeres akin to eukaryotic chromosomes. Telomeres function to seal the ends of chromosomes, stabilizing the chromosome. *Borrelia burgdorferi* telomeres consist of inverted repeat sequences and covalently closed single-stranded hairpin loops—a very different structure from that of the terminus region of circular chromosomes. As in circular genomes, chromosome replication occurs bidirectionally, but many other details including the location of the origin of replication are unknown.

Two other aspects of chromosome replication are particularly noteworthy. First, because initiation of chromosome replication can occur repeatedly prior to its completion, in circular chromosomes it has been shown that genes closer to the origin of replication are present in higher numbers than genes closer to the terminus in rapidly growing cells (Fig. 3). Hence, rapidly growing cells are partial or merodiploids, even though after growth ceases the resulting bacterial cells are haploid. Second, chromosome partitioning among the two daughter cells is a highly accurate process (loss occurs in less than 0.03% of cells) and until quite recently was thought to be a passive process coupled directly to growth of the bacterial cell prior to binary fission. Instead, chromosome partitioning appears to involve active movement, as occurs in eukaryotic cells, of the *oriC* regions toward opposite poles of the forming daughter cells.

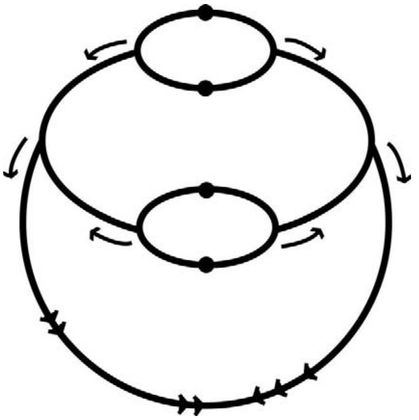


FIGURE 3 Chromosome replication in *E. coli*. Replication of circular chromosomes in bacteria can be initiated multiple times prior to completion under rapid growth conditions. The arrows indicate the direction of replication fork movement. The solid circles indicate *oriC*, and the solid triangles indicate the locations of Ter sites.

3. Plasmids

Plasmids are double-stranded DNA molecules ranging in size from ~1 kilobase (kb) to hundreds of kilobases that are physically independent of the major bacterial chromosome(s) and replicate independently. Most known plasmids are circular, although linear plasmids are found in some bacterial species. In part, the distinction between plasmids and bacterial chromosomes is historical, dating from when all bacterial cells were thought to contain a single chromosome per cell. The modern distinction is that chromosomes contain essential genes as opposed to plasmids which may contain genes expressed only under particular conditions (inducible genes). Genes carried on plasmids may code for a variety of traits, including antibiotic resistance, virulence factors, nutrient catabolism, conjugative gene transfer (see Section III,B), and anticompetitor compounds (e.g., antibiotics). Some plasmids, known as cryptic plasmids, code for no known function other than their replication. Depending on the bacterial host, genes encoded, and environmental conditions, plasmids may be parasites, commensals, or beneficial symbionts.

Plasmids can be present in single copy or up to many hundreds of copies per cell. Different plasmids use different mechanisms for replication, some with multiple origins of replication. There are two general types of replication: θ and rolling circle replication (RCR). *oriC*-type replication is an example of θ replication, in which one or two RNA primers are made for subsequent DNA replication. Unidirectional replication occurs if there is one primer and bidirectional if there are two. In RCR, a break is made in the plasmid DNA at a specific site (dso, *d*ouble-*s*trand *o*rigin) on one of the DNA strands (the + strand), which is subsequently "rolled-off" while a new strand is synthesized in its place. RCR plasmids are small and none have been found larger than 10 kb. A tendency for genetic rearrangement of RCR plasmids may be the cause for the apparent size limitation.

Regardless of the type of replication, copy number is maintained by negative feedback acting on replication initiation by one of two mechanisms. In the first, the plasmid produces a small diffusible RNA that inhibits an essential step of replication. The RNA concentration, determined by the rates of RNA production and degradation, regulates the timing of replication. In the second, a set of directed repeats (known as interons) are located within the origin of replication and serve as binding sites for a replication initiation protein. Plasmids having the same initiation control are said to be incompatible because negative feedback on copy number prevents two or more plasmids of the same initiation type from being maintained in the same bacterial cell. The degree of initiation type similarity causing incompatibility varies depending on the precise nature of control. A single base change can alleviate incompatibility for some types of diffusible RNA control, whereas interon incompatibility is more tolerant of base pair substitutions.

B. Informational Structure

Diversity among bacteria in the informational structure of the genome exists at almost all levels (e.g., GC bias and codon usage). An exception is the operon, a type of gene organization common to all bacteria.

1. GC Bias

The GC:AT (guanine + cytosine to adenine + thymine) ratio of nucleotide base composition varies greatly among bacterial genera, from 25% GC in *Mycoplasma* to 75% GC in Micrococcus, strongly contrasting with the smaller ranges of invertebrates (32–45%), vertebrates (40–45%), and plants (35–48%). Also unlike many warm-blooded vertebrates, GC base composition is relatively uniform over bacterial genomes and is not partitioned into homogeneous stretches known as isochores. The diversity of GC bias across bacterial lineages is conserved at the genus level, suggesting that base composition is generally stable over long periods of evolutionary time. There are two hypotheses for diversity of base composition across bacterial species. One hypothesis for the diversity in GC bias is selective constraint. For example, thermally stable amino acids that are encoded by GC-rich codons may be preferred in thermophilic bacteria. The second hypothesis is mutational rates biases at each of the four bases. Net GC composition results from mutations from A or T to G or C ($= u$) and from mutations from G or C to A or T ($= v$). The GC mutational pressure is the ratio $u:v$, and the expected equilibrium GC content is thus $v/(v + u)$. Direct measurements of mutational pressure cannot currently be made, but $u:v$ ratios of 3, 1, and 1/3, give the respective GC ratios of *Mycoplasma capricolum, E. coli*, and *Micrococcus luteus*.

Although the hypotheses are not mutually exclusive, there is strong evidence for the mutational hypothesis. Mutational changes are subject to selection so that base substitutions should reflect GC mutational pressure and selective constraints against mutation. This is seen in comparisons of the base composition of the first, second, and third codon positions. Most changes at the third position do not result in amino acid changes and are known as synonymous substitutions, whereas changes in the first and second positions more often cause amino acid changes. Any mutational bias should be most extreme at the third position because most substitutions do not affect the resulting protein sequence. Hence, the mutational bias of a particular bacterial species at the third position, be it AT or GC biased, is expected to exceed the average bias over the entire genome, and this is observed. Although the GC : AT ratios of *M. capricolum* and *M. luteus* over the genome are 25 and 75%, respectively, the corresponding values at the third codon position are 10 and 90%.

2. Codon Usage

The genetic code is almost completely universal across all life, with a few minor differences in some bacteria such as *Mycoplasma* and in bacterial-derived lineages such as mammalian mitochondria. More surprising is the variation in codon usage. As mentioned previously, the genetic code is very degenerate, with 61 codons coding for only 20 primary amino acids. Given the degeneracy in the genetic code, codon usage would be expected to be random for synonymous codons. However, synonymous codon usage varies across bacterial lineages, genes, and gene segments. Different bacterial lineages have highly biased codon usage, even taking into account GC bias. A large fraction of codon bias is associated with gene expression because highly expressed genes generally contain codons with abundant cognate tRNAs, which differ among species.

Differences in codon usage also exist for the initiation codon. Approximately 90% of *E. coli* genes have AUG as the ribosome start site for mRNA translation. In *E. coli* GUG or UUG can also serve as initiation sites (in this order of frequency and translational efficiency) but have a lower translational efficiency. The use of a non-AUG start codon is thought to be an adaptation to limit the production of a particular gene product in *E. coli*. For example, the use of a non-AUG start codon for adenylate cyclase limits the production of this protein, which is toxic at high levels. In contrast, approximately 30% of *B. subtilis* genes have non-AUG initiation, the order of frequency and translational efficiency is AUG > UUG > GUG, and the initiation codon has less influence on gene expression.

3. Gene Organization and Expression

At the most fundamental level, bacterial gene organization and expression is no different among bacterial species or from that of eukaryotes. All genes can be categorized into approximately three classes: protein-coding genes, RNA-specifying genes, and nontranscribed regulatory genes. Genes of the first two classes undergo transcription by a RNA polymerase, and a protein product is made from genes of the first class by rRNA translation of a mRNA transcript. However, there are numerous differences between eukaryotic and bacterial gene organization and expression. Unlike in eukaryotes, which contain multiple RNA polymerases, all bacterial RNA (mRNA, tRNA, and rRNA) are synthesized by a single RNA polymerase. Splicing of mRNA transcripts to remove introns, which are intervening noncoding regions of a gene, appears to be extremely rare in bacteria but is common is eukaryotes. The lack of a nuclear membrane in bacteria allows mRNA translation to begin prior to completion of transcription. Bacterial genes are organized into physically linked structures called operons.

a. The Operon

An operon is a transcriptional unit which minimally contains signals for DNA to RNA transcription start and stop sites but which also usually contains regulatory elements as well as RNA to protein translation-specific sites (Fig. 4). A monocistronic operon contains only a single gene, but most genes occur in polycistronic operons encoding multiple genes associated with a related function. A cistron is a DNA segment encoding a polypeptide and genes in a polycistronic operon are jointly regulated. At the most simple level, an operon has many typical elements. A repressor/activator binding site allows for the binding of a *trans*-acting factor which affects RNA polymerase binding that may itself be modulated by corepressors/inducers. The promoter region binds the RNA polymerase and a promoter-specific additional element, a σ factor, which is required for recognition and binding to the transcription initiation site. Since a single bacterium contains a variety of σ factors, global regulation of different sets of proteins can be achieved by expression of different σ factors. The most abundant σ factor found in all bacteria, σ^{70}, is essential for vegetative growth and its DNA binding is affected primarily by two DNA sequence sites, regions approxi-

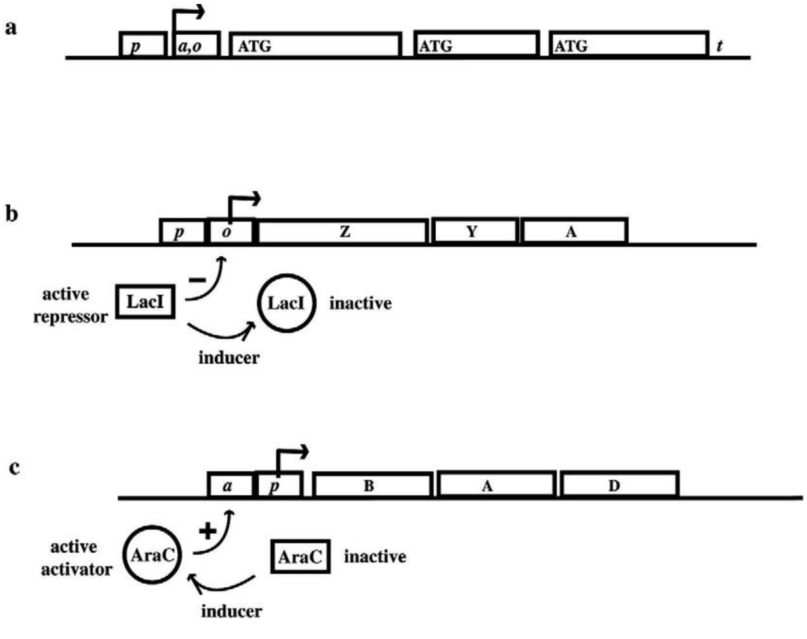


FIGURE 4 Operon structure and mechanisms of transcriptional regulation. (a) Most operons contain sites for repression (*o*, for operator binding site) and/or (*a*) activation. RNA polymerase binds at *p*, and transcription begins at a specific transcription start site (arrow). The relative position of each site varies among operons. Termination sequences (*t*) may appear after the last protein coding region. (b) Negative operon control. The genes necessary for lactose catabolism are normally repressed by the LacI protein bound to the operator. Binding of the inducer, an intermediate of lactose catabolism, to the LacI protein causes lacI to release the DNA and allows the RNA polymerase to bind and begin transcription. (c) Positive operon control. Transcription of the arabinose operon does not occur until arabinose is added and the activator (AraC) binds to the appropriate activator sites.

mately 35 and 10 bases upstream of the transcription start site. Of the more than 130 promoters in *E. coli* that have σ^{70} binding promoters, 69% are repressible, 49% are activatable, and 17% contain both repressor and activator sites.

Once the RNA polymerase has made a short piece of RNA of about 12 bases, the σ factor dissociates from the RNA polymerase and elongation of the mRNA transcript continues until terminated in one of two ways. Termination may occur by a combination of a GC-rich sequence followed closely by a run of adenines. The GC-rich region is thought to form a stem loop structure which causes the polymerase to pause and decreases the affinity of the RNA polymerase to the nascent RNA. The run of A residues results in the nascent RNA having a poly-U RNA, and A : U hybrid base pairing is relatively unstable. A second type of termination is dependent on the RNA binding protein Rho. Rho is thought to track behind RNA polymerase and terminate transcription if the polymerase pauses.

Operons encoding genes to be translated contain one or more start codons and ribosome binding sites (often known as Shine–Dalgarno sequences). Translation is initiated by the binding of the small ribosomal subunit (30S) with associated factors to the mRNA start codon and ribosomal binding site, a region approximately five bases upstream of the start codon which is complementary to the 16S rRNA component of the 30S subunit. Binding to the initiation codon by the appropriate tRNA carrying the bound cognate amino acid then leads to binding of the 50S large ribosomal subunit and formation of the bacterial 70S ribosome. Posttranscriptional regulation of expression can occur at both the ribosome binding site and the initiation codon, by the binding of translational repressors, making either or both the ribosome binding site and initiating codon inaccessible. Once begun, translation continues at an average rate of 15–20 amino acids per second until the ribosome reaches a stop codon, at which it pauses and binds a release factor which causes the completed polypeptide to be released, and the ribosome is free to dissociate from the mRNA.

b. Global Regulation

Unlike many eukaryotes, bacteria are extraordinarily proficient at altering their phenotype by gene regulation. By alteration in their patterns of gene expression, bacteria are able to take advantage of new nutrient sources, respond to harsh environmental conditions, and take alternative developmental pathways.

The classification of gene regulation in bacteria is hierarchical. Operons allow for the coordinated expression of related genes, such as those for lactose utilization (e.g., the *lac* operon). Regulons allow for coordinated expression of multiple operons in response to specific substrates, such as those for the four glycerol utilization operons in *E. coli*. In the glycerol regulon, a single repressor negatively inhibits expression of the four independent operons altering glycerol uptake and metabolism under aerobic and anaerobic conditions. Regulatory units composed of groups of operons and regulons are sometimes known as modulons. Modulon regulation is generally in response to some aspect of environmental change and the component genes typically have related tasks or activities. The aforementioned σ factors are one mechanism of global regulation, of which *E. coli* has at least six (Table II). All but one of the *E. coli* σ factors (σ^{54}) have four regions of sequence conservation. All known bacterial σ factors have sequence similarity with either σ^{54} or σ^{70}.

c. Mutational Regulation

High-frequency mutations in the genome can effectively regulate gene expression. In *E. coli,* the mutation rate per gene per generation is typically between 2×10^{-6} and 2×10^{-8}, but specific sequences (contingency loci) may have up to a 10^4 higher mutation rate. Moreover, many contingency loci usually alternate between one of two states so that the variety of outcomes is pro-

TABLE I

Main Bacterial Lineages (Phyla)

Acidobacterium
Aquifex
Bacteriodes/Cytophage
Chlamydiae
Chloroflexus
Cyanobacteria
Deinococcus
Fibrobacter
Fusobacterium
Gram-positive bacteria
Green sulfur bacteria
Leptospirillum
Planctomycetales
Protecobacteria
Spirochaeta
Synergistes
Thermodesulfobacterium
Thermophilic oxygen reducers
Verrucomicrobiales

TABLE II

Factors of *Escherichia coli*

Name[a]	Function
σ^{19}	Iron citrate transport
σ^{24}	Flagellar gene expression
σ^{32}	Temperature acclimation
σ^{38}	Nongrowth stationary phase conditions
σ^{54}	Nitrogen use
σ^{70}	Vegetative growth

[a] The numerical designation refers to the molecular weight of the σ factor.

scribed. The enhanced mutation rate increases the likelihood that at least one bacterial cell of a related population will be expressing the appropriate phenotype if environmental conditions change. Two mechanisms are known: intracellular recombination and strand slippage of DNA repeats. Recombinational regulation of cell surface components occurs in many bacteria, typically in endosymbionts avoiding host immune responses. Phase variation in *Salmonella typhimurium* is the most well-known system and is due to an invertible promoter. In one promoter orientation one surface antigen (HagA) is produced, and in the other a second antigen is produced (HagB). The promoter is flanked by a pair of repeat sequences which facilitates site-specific inversion.

Strand slippage is the insertion or deletion of nucleotides during DNA replication. This can alter the spacing between regulatory elements of an operon, enhancing or eliminating gene expression. In *Bordetella pertussis*, the bacteria causing whooping cough, a run of C nucleotides in the promoter region of a surface antigen (*fim* genes) results in frequent strand slippage and displacement of the activator binding site. Strand slippage can also result in frameshift mutations by the insertion or deletion of nucleotides other than in sets of three. All codons downstream of frameshift mutations are out of the correct reading frame, preventing expression of the protein. Runs of short DNA repeats of two, four, five, or longer greatly enhance the likelihood of strand slippage. The antigenic variation of *Neisseria gonorrhoeae* proteins essential for infection (Opa) is maintained by a five base repeat, CTTCT.

III. GENETIC EXCHANGE

Genetic exchange in bacteria was demonstrated some time ago, but it was not until quite recently that it was thought to have evolutionary importance for most species. The effects of recombination are widespread, affecting the genetic structure of bacterial populations, patterns of coevolution, and gene–gene interaction. Even so, for many species recombination is sufficiently rare that gene linkage is maintained within species. There are three mechanisms for genetic exchange among bacteria: transduction, conjugation, and transformation. All three mechanisms involve the one-way transfer of genetic material from a donor to a recipient but differ substantially in the means and control of transfer. Transduction is controlled by viral genes, conjugation by plasmids and transposons, and transformation by the bacterial chromosome.

A. Transduction

Transduction is the transfer of bacterial genetic material from one bacteria to another by a virus (a bacteriophage). Normal phage reproduction involves subversion of the host cell to produce viral-specific DNA and protein and release of infective phage containing phage DNA from the host cell. Occasionally, host cell DNA is incorporated into a phage so that the phage acts as a “vector” transferring DNA from one bacterial cell to the next. The transduced DNA can then be incorporated into the host genome by recombination. The likelihood of transduction depends on the sequence similarity between the host chromosome and the transduced DNA. Incorporation of exogenous DNA, brought into the cell by transduction or other means, drops off exponentially even with only 1% sequence divergence.

Two types of transduction are known—specialized and general—and differences in the life cycle of phage determine the possible routes for transduction to occur. Specialized transduction results when a hybrid host–viral DNA molecule is packaged into the virus, most commonly by the action of an integrative temperate phage. Temperate phages have two alternative pathways for reproduction: (i) a lysogenic pathway in which the phage genome integrates in the host genome and is replicated passively as part of the host genome and (ii) a lytic pathway in which the phage genome replicates independently and leads to the production of infective phage and lysis of the host cell. Transition from lysogenic to lytic pathways involves excision of the phage from the host DNA, and imprecise excision can result in packaging of host DNA adjacent to the site of phage DNA integration. In contrast, generalized transduction occurs as a result of errors in DNA packaging. For some phage, packaging of phage DNA is initiated at phage-specific sequences known as *pac* sites, and *pac*-like sites in the host DNA can result in its packaging into phage instead of the phage DNA. Since generalized transduction results from packaging errors, it can result from both temperate or virulent phage (phage having no lysogenic state) and can readily result in transduction of any portion of the host DNA, not just DNA adjacent to a phage DNA integration site.

B. Conjugation

Conjugation is the direct transfer of genetic material from a donor cell to a recipient cell via cell-to-cell contact. The genes transferred are usually those required for conjugation and are typically encoded on self-transmissible genetic elements such as conjugative plasmids. A variety of such elements have been identified, the most well-studied being the F plasmid of *E. coli*. Cells bearing an F plasmid (F^+ cells) extend one to three hair-like structures (sex-pili) which contact F^- cells (F plasmid-free cells) and then pull together the F^+ and F^- cells. Upon contact of the F^+ and F^- cells, DNA replication of the F plasmid is initiated and plasmid transfer from the F^+ donor to the F^- recipient begins while plasmid rolling circle replication in the donor is ongoing. If mating is not interrupted, for example, by vigorous agitation, then the entire plasmid is transferred and the recipient becomes F^+.

Genes other than those specifically for self-transmission can also be transferred by conjugative plasmids in several ways. First, R plasmids encode not only conjugative machinery but also antibiotic resistance genes. Multiple antibiotic resistances can be encoded by a single plasmid so that antibiotic resistance can quickly spread, as has been commonly observed in hospitals. Second, plasmid mobilization, the transfer of nonconjugative plasmids, can be achieved by co-occurrence of conjugative and nonconjugative plasmids in the same bacterial cell. Most plasmids appear to encode the necessary genes for mobilization even if they do not encode the complete conjugative DNA transfer apparatus. The conjugative plasmid provides those functions essential for transfer of the non-self-transmissible plasmid. Third, transfer of host genes can occur as many conjugative plasmids integrate into their bacterial host genome. Bacterial strains carrying an integrated F plasmid transfer bacterial host genes to F^- cells at a high frequency and are known as Hfr (high frequency of recombination) strains.

Although conjugative plasmids are frequently the cause of conjugation in gram-negative bacteria, conjugation in gram-positive bacteria is often due to conjugative transposons. A transposon is a transposable genetic element that is able to move directly from one DNA sequence site to other sites within the genome of a bacterial cell. At a minimum, a transposon contains only those elements necessary for it to move about the genome, but transposons may contain other genes. Even without conjugative ability, transposons may carry antibiotic genes and thus provide a rapid mechanism for movement of such genes to and from conjugative plasmids. However, some transposons are able to move from cell to cell via conjugation, although little is known about their mechanism of intercellular transfer. Gene transfer by this method is likely to be an important factor in the spread of antibiotic resistance because all known conjugative transposons characterized to date carry a tetracycline resistance gene.

C. Transformation

Genetic transformation results from uptake of free DNA from the environment and its integration into the host genome. Techniques of artificial transformation of bacteria are a cornerstone of modern molecular biology, but some bacterial genera (e.g., *Streptococcus, Neisseria, Bacillus*, and *Haemophilus*) have evolved mechanisms for frequent natural transformation. In laboratory cultures of *B. subtilis*, natural transformation and recombination begins as bacteria exhaust available resources. Cells become competent, able to pick up, bind, and internalize exogenous free DNA. There are approximately 50 DNA binding sites on the surface of a component cell and DNA binding and uptake is independent of DNA sequence and source. Once bound, DNA rapidly undergoes double-strand cleavage and DNA uptake commences ~1 or 2 min after binding. However, unlike binding and internalization, integration of exogenous DNA is severely reduced by sequence divergence. Internalized DNA that is homologous (having similar or identical base sequence with the host chromosome) can be integrated into the host chromosome. Heterologous DNA (differing in base sequence) is degraded and has no heritable effect on the recipient. The sensitivity of integration to heterologous DNA is high and is sufficient to distinguish different *Bacillus* species.

Mechanisms of transformation differ among other bacterial species. Transformation in *Streptococcus pneumoniae* appears similar to that in *B. subtilis*, except that competence is induced in response to a diffusible protein (competence factor) secreted by *S. pneumoniae* cells. In *Haemophilus influenzae*, specialized vesicle-like structures, transformosomes, extend from the cell membrane for DNA recognition and uptake. Only homologous DNA is bound to competent cells, with discrimination relying on the presence of an 11-base pair DNA sequence that occurs approximately every 4 kb in the *H. influenzae* genome.

D. Evolutionary Effects of Genetic Exchange

In the absence of genetic exchange, bacteria evolve as a series of asexual lineages. A non-recombining bacterial cell, and all the genetic material encompassed within that cell, is derived from a single immediate parental cell and all its non-recombining descendants are similarly ultimately derived exclusively from the same parental cell. A population of non-recombining bacteria initiated from a single cell is thus composed of lineages related in a dichotomously branching tree-like fashion.

The strict lineal descent of non-recombining bacteria greatly affects the dynamics and effects of adaptation. If an advantageous mutation occurs in a non-recombining population, the selective benefit conferred by the mutation affects the cell as a whole. The cells carrying the beneficial mutation can increase in frequency, displacing cells not carrying the advantageous mutation. Like the advantageous mutation, the other genes of the original mutant cell increase in frequency as well, "hitchhiking" on the selective benefit of the mutation. If the mutation fixes in the population, the complete linkage of the advantageous mutation with the rest of the genome carrying the mutation will have simultaneously have caused the genetic diversity of the population to plummet. Only those cells descended from the first cell carrying the advantageous mutation persist in the population. Hence, a series of selective sweeps by advantageous mutations, known as periodic selection, can maintain low genetic diversity within nonrecombining bacterial populations. The purging of genetic diversity will occur even if multiple adaptive mutations appear since the mutations in different cells cannot be brought together into a single cell by recombination. In addition, the rate of adaptation within a population of nonrecombining bacteria is limited because multiple advantageous mutations cannot be simultaneously fixed (Fig. 5) (de Visser *et al.*, 1999).

In contrast, multiple beneficial mutations can be fixed in sexually recombining populations and the effects of multiple selective sweeps on extant genetic variation are much reduced. Recombination eliminates the tight linkage of a beneficial mutation to the rest of the genome, with the amount of hitchhiking determined by the rate and extent of recombination. These substantial effects of genetic recombination on evolution have greatly enlivened the study of recombination in bacteria since recombination in bacteria generally appears to be very rare but of large effect.

Evidence of genetic recombination has been found in all bacterial lineages examined at the sequence level, even if recombination has not been observed. The most extensive surveys of recombination have been done in

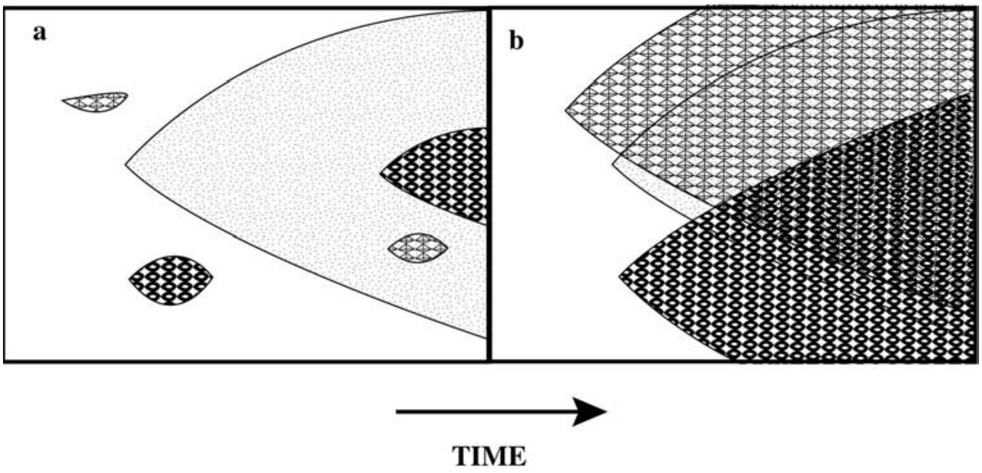


FIGURE 5 The rate of fixation of three beneficial genes in (a) an asexual and (b) a sexual population. The different patterns denote three different beneficial genes and the width of an area represents gene frequency. (a) In asexual populations, fixation of beneficial genes occurs sequentially due to the lack of recombination. (b) In sexual populations, the fixation of multiple beneficial alleles can occur simultaneously.

E. coli, a bacterial species not known to undergo natural transformation. Divergence among *E. coli* lineages is ~50 times more likely due to recombination than mutation (Guttman and Dykhuizen, 1994). As much as 15% of the *E. coli* strain K12 genome may have been acquired from bacteria differing in GC bias from *E. coli*, and such foreign DNA has accumulated at a rate approximately equivalent to that introduced by point mutations (Lawrence and Ochman, 1997). Genetic exchange is likely to have been a major source of variation in chromosome size among *E. coli* isolates, which differ by almost 20% (Bergthorsson and Ochman, 1998). Although it has often been argued that bacteria exist simply as a series of asexual lineages, with the spread of antibiotic resistance being an exception due to intense selection, genetic exchange has had widespread effects on the genetic structure of bacteria examined to date.

The effects of genetic exchange in many bacteria differ from those of obligatory sexual organisms. The combination of selection and generally far less frequent genetic recombination in bacteria maintains linkage disequilibrium, which is the occurrence of gene combinations more frequently than expected by independent assortment (Haubold *et al.*, 1998). One outcome is the evolutionary effects on operon structure. The tight linkage of the genes comprising an operon allows for the transfer of a complete metabolic function to bacteria in a single event of genetic exchange. Within a single bacterial cell, selection is relatively weak for persistence of the tight linkage of an operon. A second outcome is the potential for rapid spread of a selectively important trait, such as antibiotic resistance, even though individual recombination may be exceedingly rare. Other selectively important traits are known to have moved via genetic recombination, such as the virulence genes in enteropathogenic *E. coli*.

IV. THE EVOLUTION OF BACTERIAL DIVERSITY

The most investigated topic of biological research concerns the causes of biological diversity. Bacteria are particularly diverse and have been isolated from virtually every environment examined. The adaptive potential of bacteria to take advantage of diverse environments is due to at least four reasons. First, bacteria are small. This allows them to take advantage of microenvironments arising from spatial heterogeneity—environments that are abundant but incapable of supporting metazoan populations. Such environments include different animal and plant hosts. Adaptation to microenvironments can be readily observed in laboratory cultures of bacteria in as little as 3 days (Rainey and Travisano, 1998). Their small size also allows them to grow to large populations (10^8 cells or more).

Second, environment-specific adaptation can occur readily since genetic recombination is too infrequent to impede adaptation. Frequent recombination could prevent specialization to different niches as the genes involved in specialization to different niches are mixed with one another by recombination. However, even though genetic exchange of selective important genes can readily be demonstrated, ecological specialization can rapidly evolve. This has been observed in laboratory populations of bacteria several times (Rosenzweig *et al.*, 1994) but can also be seen in nature (e.g., antibiotic resistance).

Third, bacteria are phenotypically plastic, allowing them to take advantage of temporally varying environments. Bacteria can drastically alter their requirements for growth, including changes from heterotrophy to prototropy and from aerobic to anaerobic growth. Many bacteria are able to grow despite rapid changes in osmolarity, pH, and temperature, and they can withstand greater changes by ceasing growth. If conditions are sufficiently poor, many can enter a dormant metabolic state, such as a spore, that can remain viable for decades or longer without water or nutrients.

Finally, as mentioned at the beginning of the article, bacterial phenotypes are evolutionarily plastic and adaptation can be finely tuned to environmental conditions. For example, *E. coli* grown in a spatially and temporally constant environment for 800 generations will diverge from a single lineage into three distinct lineages when provided with glucose as a limiting resource. The dominant lineage is a glucose specialist that excretes acetate and glycerol as waste products, which respectively support an acetate specialist and a glycerol specialist. By optimizing different steps in the catabolism of a single limiting resource, otherwise identical bacterial genotypes can simultaneously persist (Rosenzweig *et al.*, 1994). A second example is the ability of bacteria to adapt to even extremely harsh conditions by specialization. *Thermus aquaticus* can grow at temperatures up to 102°C but not less than 65°C, whereas *Vibrio marinus* can grow at −14°C but not above 25°C. However, despite radically different selective conditions, the genetic systems of harsh environment specialists show little difference from those of other bacteria. The only apparent shortcoming of the genetic architecture of bacteria is the lack of true bacterial multicellularity.

See Also the Following Articles

ARCHAEA, ORIGIN OF • BACTERIAL BIODIVERSITY • EUKARYOTES, ORIGIN OF • NUCLEIC ACID BIODIVERSITY

Bibliography

Atlas, R. M., and Bartha, R. (199). *Microbial Ecology: Fundamentals and Application*, 4th ed. Benjamin/Cummings, Menlo Park, CA.

Bergthorsson, U., and Ochman, H. (1998). Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Mol. Biol. Evol.* **15**, 6.

de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L., and Lenski, R. E. (1999). The fate of competing beneficial mutations in asexual populations. *Science* **283**, 404.

Dykhuizen, D. E. (1998). Santa Rosalia revisited: Why are there so many species of bacteria? *Antione van Leeuuwenhoek* **73**, 25.

Guttman, D. S., and Dykhuizen, D. E. (1994). Clonal divergence in *Escherichia coli* is a result of recombination, not mutation. *Science* **266**, 1380.

Haubold, B., Travisano, M., Rainey, P. B., and Hudson, R. R. (1998). Detecting linkage disequilibrium in bacterial populations. *Genetics* **150**, 1341.

Lawrence, J. G., and Ochman, H. (1997). Amelioration of bacterial genomes: Rates of change and exchange. *J. Mol. Evol.* **44**, 383.

Lengeler, J. W., Drews, D., and Schlegel, H. G. (Eds.) (1999). *Biology of the Prokaryotes*. Blackwell, Stuttgart.

Li, W.-H. (1997). *Molecular Evolution*. Sinauer, Sunderland, MA.

Neidhardt, F. C. (Ed.-in-chief) (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. ASM Press, Washington, DC.

Rainey, P. B., and Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature* **394**, 69.

Rosenzwieg, R. F., Sharp, R. R., Treves, D. S., and Adams, J. (1994). Microbial evolution in a simple unstructured environment: Genetic differentiation in *Escherichia coli*. *Genetics* **137**, 903.

Salyers, A. A., and Whitt, D. D. (1994). *Bacterial Pathogenesis: A Molecular Approach*. ASM Press, Washington, DC.

Sherris, J. C. (Ed.) (1990). *Medical Microbiology: An Introduction to Infectious Diseases*. Appleton & Lange, Norwalk, CT.

Sonenshein, A. L. (Ed.-in-chief) (1993). *Bacillus subtilis and Other Gram Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics*. ASM Press, Washington, DC.

BEETLES

Henry A. Hespenheide
University of California, Los Angeles

GLOSSARY

fauna The animal species living in a defined area.
leaf beetles Members of the family Chrysomelidae.
phytophagous Feeding on plants; also, herbivorous.
weevils Members of the beetle family Curculionidae.

THE DIVERSITY OF BEETLES as used here refers principally to the number of species of beetles; that is, a "large" group has more species than a "small" group. This use of "diversity" reflects most scientific discussions of biodiversity, though it may also include variation in morphology, ecology, or genetics, or consideration of numbers of genera or families or taxa at other levels of classification. In this review, the argument is made that diversity in numbers of species is partly a consequence of the diversity of ecological types, so that these are in fact not fully independent.

I. INTRODUCTION

In terms of numbers of species, arthropods are the largest phylum of living organisms, insects are the largest group within the arthropods, and beetles are the largest order among the insects. Beetles presently comprise about a quarter of all described, extant organisms. This numerical dominance led the British biologist J. B. Haldane to answer the question of what one might learn about the mind of the Creator from a study of his creation with the famous reply: "An inordinate fondness for beetles."

In 1982, T. L. Erwin used samples of beetles to estimate that the number of species of living organisms may be 30 million or more. The ensuing controversy (especially Stork, 1988) rekindled interest in the more general question of the magnitude of global biodiversity (May, 1988) and how to measure it. These discussions did not question the usual assumption that beetles represent the largest part of global biodiversity, although mites and nematodes may in fact be more diverse than beetles. More recently and narrowly, a proposal to "explain" the diversity of beetles has itself engendered controversy. Both questions are reviewed here.

II. WHAT BEETLES ARE

The order Coleoptera is usually divided into four suborders and 160–170 families depending on the author. The suborders Archostemata and Myxophaga each consist of four families and fewer than 100 extant species. The Adephaga includes about a dozen families, with most of the species in the large family of ground beetles Carabidae. All other beetles—150 families and 90% of the 340,000 species—are classified as Polyphaga. Somewhat more than two-thirds of the species belong to the eight largest families discussed here, each with more than 10,000 species, but there are 30 other families that have between 1000 and 10,000 species (Table I). The most thorough review of the biology of members of the order is found in Crowson (1981).

Because they are well sclerotized, there is a good fossil record for beetles. Beetle-like insects first appeared in the Lower Permian, modern families are known from the Cretaceous, and extant genera date from the Oligocene and Miocene. Ecologically, beetles are the most diverse order of insects, which is a primary factor in their great species richness. There are large families of beetles that live in fresh-water habitats, that are predaceous and parasitic, and that feed on fungi and detritus, but the largest proportion of described species feed in some way on plants. Among these phytophagous beetles, the largest number of species are wood-borers; a much smaller number feed directly on leaves, and some feed on seeds.

III. DIVERSITY—HOW MANY BEETLES ARE THERE?

Gaston (1991) has reviewed current estimates of the number of insect species described, favoring those of J. F. Lawrence for beetles. Lawrence (1982) put the numbers of beetles at 340,500 species divided into about 25,000 genera and placed in 169 families. His estimate of the number of species falls in the middle of estimates (290,000–370,000) cited by Gaston, and roughly equal to the sum of the number of species in the three next largest insect orders (Diptera, 98,000–120,000 species; Lepidoptera, 112,000–165,000 species; Hymenoptera, 100,000–130,000 species). Estimates of the total number of living species, described and undescribed, are difficult to relate to the number of described species for two major reasons. First, most taxonomic research has been carried out in the north temperate regions of the world on organisms that live there, whereas most insect species are tropical. Second, different groups of insects have received differential amounts of study from scientists, both within the beetles and between beetles and the other insect orders. This article briefly reviews what is known about the major beetle families, considers explanations for the success of the order, reviews a few ecological and biogeographic patterns in beetle diversity, and then discusses the problems in estimating what we don't yet know about beetle diversity.

As stated earlier, eight families of beetles currently account for more than half of all beetle species. However, unequal efforts in the collection and study of different groups very likely mean that the relative sizes of these families will change when the less well studied taxa become better known. Some of the better-studied groups have already begun to be used to test hypotheses of mechanisms that might promote or maintain beetle diversity; for example, whether speciation in plant-feeding beetles parallels speciation in their hosts. The eight major families are considered in order of their traditional phylogenetic placement.

1. Carabidae

The ground beetles (30,000 described species, including the tiger beetles, sometimes placed in the separate family Cicindelidae) are primarily predatory. Although some carabid groups are very rich in tropical regions, others do not show as great an increase in diversity there as do other beetles. Despite the vernacular family name, tropical forms are commonly arboreal. Tiger beetles are well studied taxonomically, have been used to test ecological and evolutionary hypotheses, and have been suggested as indicator species for biodiversity studies (Pearson and Cassola, 1992).

2. Staphylinidae

The rove beetles (30,000 species) are both numerous in species and diverse ecologically, including predatory, fungus, and detritus feeders. The staphylinids are the least well known family of beetles, with perhaps only 10% of extant species described, and may eventually prove to be the largest family. They are abundant in canopy fogging samples in tropical regions.

3. Scarabaeidae

The scarabs (25,000 species) are primarily phytophagous, but with one group well known as dung beetles. Because they are popular with collectors, they are among the best known of the beetle groups. The actual number of extant species is unlikely to be more than twice as many as those already described.

4. Buprestidae

The jewel beetles (15,000 species) include a majority of wood-boring species and several groups of leaf-miners. Although larger species are popular with collectors, the

TABLE I

Diets and Numbers of Species and Genera of the Larger Families of Beetles (More Than 1000 Species)[a]

Taxon	Genera	Species	Diet[b]							
			A	P	D	F	Ph	W	S	L
Adephaga										
Caraboidea										
Carabidae	**1500**	**30,000**		P						
Dytiscidae	120	3000	A							
Polyphaga										
Hydrophiloidea										
Hydrophilidae	125	2000	A							
Histeroidea										
Histeridae	200	3000		P						
Staphylinoidea										
Liodidae	250	2000				F				
Scydmaenidae	75	2000			D					
Staphylinidae	**1500**	**30,000**		P	D	F				
Pselaphidae	650	5000				F				
Scarabaeoidea										
Lucanidae	100	1200					Ph			
Scarabaeidae	**2000**	**25,000**					Ph			
Buprestoidea										
Buprestidae	**400**	**15,000**						W		
Elateroidea										
Elateridae	400	9000					Ph	W		
Eucnemidae	190	1200						W		
Cantharoidea										
Lycidae	150	3500		P						
Lampyridae	100	2000		P						
Cantharidae	135	5000		P						
Bostrychoidea										
Anobiidae	180	2150						W		
Cleroidea										
Cleridae	150	4000		P						
Melyridae	200	5000		P						
Cucujoidea										
Nitidulidae	160	3000			D					
Cucujidae	75	1200		P						
Erotylidae	30	2500				F				
Endomychidae	120	1300				F				
Coccinellidae	500	4500		P						L
Colydiidae	180	1300		P						
Mordellidae	100	1200					Ph	W		
Oedimeridae	100	1000						W		
Anthicidae	100	3000					Ph			
Meloidae	120	3000		P						
Tenebronidae	**1700**	**18,000**					Ph			
Chrysomeloidea										
Cerambycidae	**4000**	**35,000**						W		
Bruchidae	60	1500							S	
Chrysomelidae	**2500**	**35,000**								L
Curculionoidea										
Anthribidae	325	2600				F				
Attelabidae	100	2100								L
Apionidae	26	2200					Ph			
Brenthidae	325	2300						W		
Curculionidae	**4500**	**50,000**					Ph	W		

[a] Data from Lawrence (1982). Families listed in boldface account for more than two-thirds of all beetle species.

[b] A = aquatic, D = detritus, F = fungi, L = leaves, P = predatory, Ph = phytophagous, S = seeds, W = wood-boring.

smaller species are primarily subtropical and tropical and very poorly known. The genus *Agrilus* may be the largest genus of living organisms with 2000 described species and many more undescribed—there are nearly 350 undescribed species among more than 600 known from México alone. At La Selva Biological Station in Costa Rica, 24 of 29 *Agrilus* (83%) are undescribed; of 206 Buprestidae, 147 (71%) are leaf-miners and 62% are undescribed. The family should more than double in size when completely known.

5. Tenebrionidae

The darkling beetles (18,000 species) are phytophagous and have been relatively popular and well studied in temperate and subtropical regions. I have seen no discussion of its current taxonomic status or estimates of undescribed species.

6. Cerambycidae

The long-horned beetles (35,000 species) are wood-borers, very popular among collectors, and well studied. Of 274 species collected at La Selva, only 23 (8%) are undescribed, although another 28 (10%) are still undetermined. It is possible that the number of species will double when the rich South American fauna is more fully known. This and the following two families constitute the Phytophaga and are the largest beetle families in described species.

7. Chrysomelidae

The leaf beetles (35,000 species) are by far the largest group of beetles that primarily feed on the leaves of plants. Some groups within the family are relatively well studied, even within tropical areas (Cassidinae), whereas others are not (Alticinae). Members of the family are common in canopy samples and have been shown to speciate in parallel with their plant hosts. The success of beetles overall has been attributed to the success of this and the other two families in the Phytophaga. When the total chrysomelid fauna is known, it will probably double the number of currently known species.

8. Curculionidae

The weevils (50,000 species) are the largest family of living organisms. Although they are primarily phytophagous and associated with angiosperm plants in a variety of ways, many (and perhaps most) are wood-borers (many Cryptorhynchini, Molytini, and Zygopini) rather than leaf-eaters, in some cases cultivating fungi in their tunnels (Platypodinae and some Scolytinae). The great majority of weevils are endophytic; the few that feed directly on leaves often do so as leaf-miners. Sampling at tropical localities has yielded large numbers of weevils. At Barro Colorado Island in Panamá, three years of light trapping (Wolda *et al.*, 1998) yielded 1240 species of Curculionoidea (excluding Platypodinae and Scolytinae, but including about 200 species from four other families). Sampling at La Selva has collected well over a thousand species of weevils, including 504 in the largest tribe, Zygopini, of which 425 (84%) are undescribed. The number of species is probably two to three times those presently known.

IV. EXPLANATIONS FOR THE SPECIES RICHNESS OF BEETLES

Farrell (1998) recently claimed to have verified an earlier proposal that the great species richness of Coleoptera is due to the association of the largest lineages within the Phytophaga (Cerambycidae, Chrysomelidae, Curculionidae) with angiosperm plants. Indeed, beetles feed on angiosperms in a number of ways. The narrowest form of phytophagy is to feed on leaves (foliivory), but very few beetles do that, primarily Chrysomelidae and leaf-miners in several families. As foliivores, the Lepidoptera are much more successful (most of 140,000 species). A few phytophagous beetles feed on seeds (Bruchidae, some Curculionidae), but most are wood-borers (Buprestidae, Cerambycidae, many Curculionidae). Perhaps more than qualities of the host itself, it is the ability of beetles to feed endophagously in wood and other plant tissues that makes them so successful. Anderson (1995) has attributed the success of weevils, the most species-rich lineage, to the evolution of the rostrum as an aid to oviposition in plant tissues. It is more difficult for Diptera and Lepidoptera to feed endophagously because adults lack chewing mouthparts with which to exit plant tissues.

Another problem with the Farrell hypothesis is that even if the Phytophaga were removed from the beetles, the remaining taxa are still twice as numerous as any other insect order. If the Staphylinidae are as species rich as has been suggested (300,000 species), that family alone would be more than twice as large as the currently known Phytophaga and about the same size as estimates of total Phytophaga. The success of beetles is almost certainly due to their use of so many major resources, not just their intensive use of one of them. Of the 8 largest beetle families, the Carabidae are primarily predaceous and the Staphylinidae are partly or largely so; 10 of the other 30 families with more than 1000 species are also primarily predaceous. Five of the larger families and some Staphylinidae feed on fungi, which is another major adaptive zone for beetles. Aquatic habitats are used by two large families of beetles and only by Diptera among the other three major orders.

Ecological samples of canopy insects, although dominated by phytophages, include significant portions of species of other feeding types. If only half of all beetle species feed on angiosperm plants, it could be argued that beetles are in fact *less* successful than expected, for the reason that angiosperms constitute a much larger resource in terms of productivity and biomass than the other trophic levels in which beetles feed with equal success. In sum, there is no single simple explanation for the great diversity of beetles. Although association with angiosperms has obviously been a major factor in the diversification of beetles, it has been so in a very complex way that may be due less to the plants as a resource and more to the features of the beetles ("preadaptations") that allowed them to use the plants endophagously.

V. TAXONOMIC AND ECOLOGICAL PATTERNS

Most large taxa have greater numbers of species in tropical regions than in temperate ones, and that pattern is likely to hold for families of beetles. Increase in species numbers at lower latitudes is not usually a matter of uniform increases in all genera within the taxon, but rather of changes in dominance among taxa. For example, in the Buprestidae, wood-boring genera dominate in temperate and subtropical regions both north and south of the equator, but leaf-mining genera dominate near the equator. Among the leaf-miners, the genera *Brachys* and *Taphrocerus* dominate in temperate and subtropical regions, but the genera *Hylaeogena, Leiopleura,* and *Pachyschelus* dominate near the equator. In the Chrysomelidae, the subfamily Clytrinae has a peak in species richness in the subtropics, as does the zygopine weevil genus *Cylindrocopturus.*

More complex relationships between species diversity and ecological diversity are suggested by some data. The wood-boring Cerambycidae increase in absolute species richness as one moves towards the tropics, but F. T. Hovore (unpublished data) has found that species richness relative to potential woody plant host species diversity actually decreases at lower latitudes because of a greater increase in the number of woody plant taxa (Fig. 1). This is not unexpected in that many tropical plant species are rare, and Southwood and others have shown that rarer plants have fewer host-specific herbivores than do more common plants. In the tropical Cerambycidae, many of the species are associated with a few larger, more abundant host plant genera such as *Inga* and *Ficus,* and host generalists may be common.

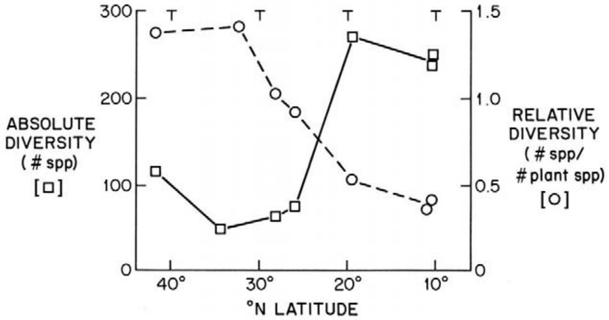


FIGURE 1 Diversity of the beetle family Cerambycidae at several sites as a function of latitude. Number of species is given on the left axis (squares)—the number increases from the temperate zone to the tropics. Relative diversity is shown on the right axis, expressed as the number of species of beetles compared to the number of species of potential woody plant hosts (circles)—the number of beetle species increases less rapidly than the number of potential hosts. (Data from F. T. Hovore, unpublished.)

Mimicry is more common in tropical beetle species than in temperate members of the same taxon. The proportion of species participating in mimicry complexes has been shown to increase with decreasing latitude in *Agrilus* (Buprestidae) mimicking the ant genus *Zacryptocerus,* and in zygopine weevils mimicking a variety of models. On the other hand, mimicry of and by clytrine chrysomelids is most frequent in the subtropical regions of México, where the clytrines are most species rich.

VI. HOW MUCH DON'T WE KNOW?—LOCAL FAUNAS

Estimates of how many beetle species exist in various families and the contribution of beetles to global biodiversity should be placed in the context of our extensive, basic ignorance of the ecology and distribution of beetles at local and regional scales. Except perhaps for England, there is no place on earth of any size and ecological diversity where one can make a complete list of all beetle species. In areas where checklists have been attempted, the larval biology and population ecology of many or most species are completely unknown. Lists of described species are incomplete, and a significant number of species are undescribed in all but the most popularly collected taxa. A few small taxa have been surveyed both intensively at a few sites and more or less extensively at scattered sites (Pearson and Cassola, 1992). In general, recent extrapolations to global diversity estimates have been made from samples of local faunas.

Beetles occupy such a wide variety of ecological roles in communities that no single method of sampling is adequate, even when the taxonomic scope of the beetles studied is relatively narrow. In recent years, Malaise trapping and canopy fogging have become popular, but flight intercept traps, sweep netting, blacklighting, Berlese sampling, pitfall traps, baiting, beating, systematic rearing, and other methods have been used as well. Malaise traps sample active, flying insects (e.g., Mordellidae), and beetles in Malaise samples have been used as indicators of specific habitats. Light trapping (Wolda *et al.,* 1998) is successful in sampling cryptorhynchine and molytine weevils, but is poor for diurnally active zygopines and baridines. Canopy fogging has been widely used in recent years to sample beetle faunas.

Several major projects have attempted to inventory the arthropod faunas of single, diverse localities. The Arthropods of La Selva (ALAS) project in Costa Rica, for example, uses a combination of four standardized sampling techniques (Malaise trapping, canopy fogging, litter sampling, and light trapping), as well as group-specific collecting methods, to survey a variety of arthropods including several beetle groups. A similar variety of sampling techniques were employed in Sulawesi (Hammond, 1993). These and other such projects in tropical regions have been slow in producing more than partial faunal lists because of the massive taxonomic problems associated with rich tropical faunas that have been largely unstudied (see the following sections). The current ecological concept of metapopulations raises questions about the meaning of local faunas by suggesting that local diversity is a dynamic function of regional diversity. The idea that some "sink" populations are maintained by regular immigration from "source" populations elsewhere may explain anecdotes of the sudden and irregular appearance and disappearance of species of leaf-mining Buprestidae at La Selva in Costa Rica.

VII. ESTIMATING DIVERSITY GLOBALLY

The problem of rare species has both ecological and taxonomic–evolutionary implications. Rare species are difficult to sample and difficult to interpret when they are sampled. Apparent rarity can be due to several factors, including methodological artifacts, disturbance by "tourists," seasonal phenology, cyclic populations, or true biological rarity. Furthermore, small populations may be characteristic of tropical species. Three years of light trapping by Wolda at Barro Colorado Island, Panamá, yielded over 95,000 weevil specimens, but 28% of the 1239 species were represented by a single specimen. In smaller samples at six other localities in the same study, unique specimens represented 39–51% of the species. Canopy fogging samples show comparably high proportions of rare species. Reliable taxonomic decisions about what is or is not a valid biological species often require series of specimens to interpret variation; adequate series for studies are even more necessary when specimens from several localities are involved. Many described tropical species are known only from unique type specimens. The question of what is a species has been justifiably raised in discussions of biodiversity at regional and global geographic scales.

Erwin (1982) startled the scientific world by calculating that 1200 beetle species (955 counted, 206 weevils estimated, then "rounded up" to 1200) sampled from 19 individual trees of *Luehea seemannii* implied that there were 30 million species of insects worldwide. His extrapolation depended on three important quantities: his estimate of the proportion of host-specific beetles, the fraction found exclusively in the canopy, and the assumption that all tree species have the same number of herbivores. The last of these is certainly not true—rare tree species have fewer host-specific insects (see earlier discussion).

Empirical studies that sample both the canopy and lower levels in the forest have not shown a large and/or distinct canopy fauna. Gaston (1991) interviewed taxonomic specialists of many insect taxa, including for beetles, and concluded that the proportions of undescribed species in tropical samples were high but not high enough to warrant an estimate of 30 million species (if 1 million species are known and 30 million exist, undescribed species would comprise 97% of samples).

Other means of estimating global species richness have been used. Rates of description of new species of beetles ("trend lines") have been extrapolated to an asymptote, but this method incorrectly assumes constant levels of scientific study and publication. May (1988) compared observed with theoretical distributions of body size and abundance to estimate the numbers of undescribed species, primarily those in the smaller size classes. It is certainly true that larger organisms are more likely to be collected and described than are smaller organisms. In each of the other major orders of insects, it is the smaller species that are least well known: for Lepidoptera, it is the microlepidoptera; for Hymenoptera, the smaller parasitic wasps in the Chalcidoidea; for Diptera, the small Nematocera and Phoridae. In beetles, Gaston showed that larger beetles were described earlier and that the average size of the species being described decreases as one nears the present. Of the zygopine weevil species identified at La Selva, 57% of weevils larger than 4 mm in length are undescribed (88 measured), but 81% of those less than 4 mm are undescribed (146 measured). Methods of estimating unsampled species from theoretical abundance frequency distributions depend on the mode of the distributions being defined, which is not usually possible for tropical samples. It seems to be true for all estimates of global beetle or insect species numbers, regardless of the sampling method used, that what is not known is so large compared to what is known that a precise estimate is impossible.

VIII. WHAT DO WE KNOW AND HOW LIKELY ARE WE TO EVER KNOW?

Beetles are a very large group of organisms and are diverse both taxonomically and ecologically. The 169 families and 340,000 described species currently comprise somewhat less than half of all known insects and about a quarter of all living organisms. About 40 families of beetles have more than 1000 described species, and 8 families have more than 10,000 species. About half of all beetle species feed on plants, usually as endophages on woody or other support tissue, but there are a large number of predatory species, as well as lesser but significant numbers of fungivores, detritivores, and aquatic forms. This ecological diversity precludes simple explanations for the diversity of beetles. However, many beetle families show general patterns of distribution, for example, greater diversity in tropical latitudes and the greater importance there of antipredator defenses such as mimicry.

These crude generalizations aside, we actually know rather little about most beetles. Because of the popularity of certain beetle groups, the various families have been unequally studied. Parts or all of the families Carabidae, Scarabaeidae, Buprestidae, Tenebrionidae, and Cerambycidae have been relatively better studied, whereas the three largest families (Staphylinidae, Chrysomelidae, and Curculionidae) are the most poorly studied. Ironically, as we have become aware of the existence of large numbers of undescribed species by using new sampling methods in previously little-collected tropical areas, the number of professional taxonomists able to name and describe species has been declining. Many positions for taxonomists have been lost at most major museums, funding for taxonomic research has declined, and the trend in research has been toward genetic analyses for the purpose of elucidating phylogenetic relationships rather than describing species. Money has become available for sampling and conducting species inventories, especially in threatened tropical areas, but not for processing and studying the material collected beyond simple sorting of a few target or focal taxa and counting of their putative "morphospecies." May (1988) showed that, although beetles are the largest group of organisms, they have the lowest rate of publications per described species of all organisms except nematodes. He expressed amazement that there is, in fact, no complete worldwide catalog of insects that would allow one to know how many are actually described (hence the varying "estimates" of even the numbers of described species).

At the level of local or regional faunas, there are few sites outside of Western Europe where beetle faunas are more than partially known. The richest areas—the tropics, and especially the Neotropics—are virtually unknown faunistically outside of samples taken at a few widely scattered sites. Many large regions of the tropics have never been significantly collected for any beetle groups. Costa Rica's National Biodiversity Institute (INBio), including the ALAS project at La Selva, has made the most extensive collections toward a regional fauna of any tropical country, but their collections favor larger species, are only partially mounted and sorted, and are only beginning to be studied. Overall, then, there is no good estimate of global species richness for beetles and few data on which to base such an estimate.

As the magnitude of how much we still don't know becomes clearer, there are serious questions about whether we will ever understand the extent of beetle diversity in light of the declining number of taxonomically trained scientists and the possibility that many species will go extinct because of human activities before they are ever collected or studied. The decline in the number of scientists who are able or willing to describe species is unlikely to be halted. Furthermore, widespread beetle extinction over the next century is especially likely in the face of the accelerating environmental destruction of the richest sites in the tropics to accommodate growing human populations and, even more significantly, to maintain Western European and North American standards of living. Unfortunately, it is a safe guess that we will never completely know how many beetles there were.

See Also the Following Articles

ARTHROPODS, AMAZONIAN • INSECTS, OVERVIEW • INVERTEBRATES, TERRESTRIAL, OVERVIEW

Acknowledgments

I thank F. T. Hovore for reviewing this manuscript and allowing me to publish some of his data. I have been supported by the Research Committee of the UCLA Academic Senate and by grants in support of the ALAS Project (National Science Foundation Grants BSR 9025024, DEB 9401069, and DEB 9706976).

Bibliography

Anderson, R. S. (1995). An evolutionary perspective on diversity in Curculionoidea. *Mem. Entomol. Soc. Washington* **14**, 103–114.

Crowson, R. A. (1981). *The Biology of the Coleoptera.* Academic Press, London.

Erwin, T. L. (1982). Tropical forests: Their richness in Coleoptera and other arthropod species. *Coleopts. Bull.* **36**, 74–75.

Farrell, B. D. (1998). “Inordinate fondness” explained: Why are there so many beetles? *Science* **281**, 555–559.

Gaston, K. J. (1991). The magnitude of global insect species richness. *Conservation Biol.* **5**, 283–296.

Hammond, P. M. (1993). Insect abundance and diversity in the Dumoga-Bone National Park, N. Sulawesi, with special reference to the beetle fauna of lowland forest in the Toraut region. In *Insects and the Rain Forests of South East Asia* (*Wallacea*) (W. J. Knight and J. D. Holloway, eds.), pp. 197–254. Symposium of the Royal Entomology Society of London, Royal Entomology Society of London, London.

Hespenheide, H. A. (1994). An overview of faunal studies. In *La Selva: Ecology and History of a Neotropical Rainforest* (L. McDade, K. S. Bawa, H. A. Hespenheide, and G. S. Hartshorn, eds.), pp. 238–243. University of Chicago Press, Chicago.

Lawrence, J. F. (1982). Coleoptera. In *Synopsis and Classification of Living Organisms,* pp. 482–553. McGraw-Hill, New York.

May, R. M. (1988). How many species are there on earth? *Science* **241**, 1441–1449.

Pearson, D. L., and F. Cassola. (1992). World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): Indicator taxon for biodiversity and conservation studies. *Conservation Biol.* **6**, 376–391.

Wolda, H., C. W. O’Brien, and H. P. Stockwell. (1998). Weevil diversity and seasonality in tropical Panama as deduced from light-trap catches. *Smithsonian Contr. Zool.* **590**, 1–79.

BIODIVERSITY AS A COMMODITY

Geoffrey Heal
Columbia University and Stanford University

GLOSSARY

first theorem of welfare economics If all goods are private and all private and social costs are equal, then an economy with a complete set of competitive markets operates in a way that is Pareto efficient.

insurance A contract under which a person or institution buys the right to be compensated in the event of a specified harmful occurrence. The payment, referred to as the insurance premium, is paid before it is known whether the harmful event will occur.

intellectual property rights The rights of developers of ideas and techniques to require payment for their use by others and to prevent their use by others unless such payment has been made.

invisible hand The term used by Adam Smith to describe the capacity of a decentralized market system to attain efficiency.

pareto efficient A pattern of operation of an economy is said to be Pareto efficient if it is impossible to change its operation so that everyone gains or at least someone gains and no one loses.

public good A good whose consumption is nonrival and nonexcludable. Nonrival means that one person's consuming it does not preclude another from doing likewise. Nonexcludable means that the provider of the good cannot ensure that only those who have paid can benefit from its provision. Knowledge is a public good: My knowing something does not conflict with your knowing the same fact, and those who develop knowledge cannot ensure that only people and institutions that have contributed to the costs can benefit. An apple, in contrast, is a private good: If I eat it, you cannot. Also, apple producers can ensure that only those who pay for them can eat them.

social costs The total costs to society of an action. These may exceed (or in some cases be less than) the private costs, which are the costs of that action to the individual or institution executing it.

WHAT IS THE NATURE OF BIODIVERSITY as an economic commodity and why does it matter? How would its conservation contribute economically to our well being? I consider three issues: Why is biodiversity important from an economic perspective? What kind of commodity is it? and Does our usual economic mechanism, the market system, have the capacity to appreciate the economic value of biodiversity? I first characterize biodiversity from an economic perspective, and then

Encyclopedia of Biodiversity, Volume 1

consider the capacity of our main economic institutions to realize the value of biodiversity and ensure that it is treated in a way commensurate with its importance.

Many distinguished scientists have argued strenuously for the conservation of biodiversity as a significant human priority (Ehrlich, 1988; Raven and Williams, 1999; Wilson, 1988). Because economic activity is the main driver of biodiversity loss, such conservation would undoubtedly have important economic implications. It would require economic changes and would certainly be associated with economic costs, although these could be more than offset by the gains. The proposal implicit in these arguments is in effect that we—society—should "buy" biodiversity by changing our economic modus operandi and incurring conservation costs.

What are we buying if we make such a deal? What is the nature of biodiversity as an economic commodity and why does it matter? How would its conservation contribute economically to our well being? These are the themes I address in this article. I consider three issues:

What are the economic functions of biodiversity?
What kind of commodity is biodiversity?
Does our usual economic mechanism, the market system, have the capacity to appreciate the economic value of biodiversity?

In other words, I first characterize biodiversity from an economic perspective, and then consider the capacity of our main economic institutions to realize the value of biodiversity and ensure that it is treated in a way commensurate with its importance. With regard to the first of these tasks, assessing why biodiversity is important economically, I draw extensively on recent literature in ecology, which I do with some trepidation because this is not my field of expertise.[1] However, this is unavoidable: A serious analysis of the economic contributions of biodiversity has to draw on scientific understanding of how diversity contributes to the functioning of the natural environments that provide crucial infrastructure to human societies.

I begin with an attempt to explain what biodiversity contributes, in economic terms, to human societies. Why is biodiversity important? The reasons can be classified under the following headings. Biodiversity provides or enhances

Ecosystem productivity
Insurance
Knowledge
Ecosystem services

There is some overlap between these concepts, but nevertheless they are helpful as a guide to thinking through the issues. All of them are economically important categories.

I. BIODIVERSITY AND PRODUCTIVITY

How does biodiversity contribute to productivity? There is experimental evidence that plant systems with more biodiversity are on average more productive than those with less biodiversity. A good illustration of this is work done by David Tilman at the University of Minnesota (Tilman and Downing, 1994; Tilman *et al.*, 1996, 1997). He planted many similar plots of land with a variety of grassland plants, some with many species and some with a much smaller number. Each plot was planted with the same mix year after year, and several indicators of plot performance were recorded, including the amount of biomass grown and the proportion of the nutrients available that were taken up by the plants. Biomass refers to the total dry body weight of the plants: It is a measure of the amount of carbon from the atmosphere that is photosynthesized into carbohydrate. Tilman and others performing similar studies found that, on average, during a period of about 20 years, plots with a more diverse collection of species performed better than those with a less diverse collection.

What does more or less diverse mean here? There are two dimensions to diversity: diversity of functional groups, or of plant types, and diversity of plant species within a functional group. Plants are classified into functional groups on the basis of their intrinsic physiological and morphological characteristics, such as whether they fix nitrogen, have three carbon or four carbon photosynthetic pathways, or are woody (Tilman *et al.*, 1997). These characteristics influence the plant's resource requirements, seasonality, and life history. A key aspect of diversity is measured by the number of different functional groups represented by the plants on the plot. This is called, not surprisingly, functional diversity. Species diversity refers to the number of different species within each functional group, or to the total number of species present. This latter measure is sometimes called diversity per se. Clearly, there is a correlation between diversity per se, the total number

---

[1] My understanding of the relevant issues has benefited from conversations with Gretchen Daily, Paul Ehrlich, David Tilman, and Peter Vitousek.

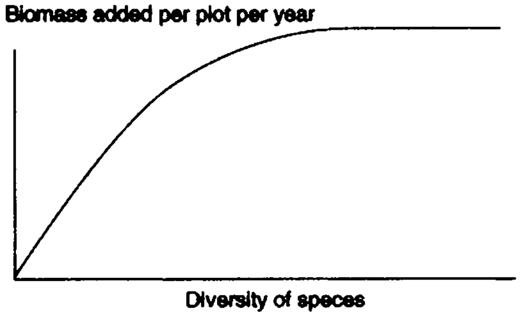


FIGURE 1 Average amount of biomass grown per year on a plot of a given size.

of different species present, and functional diversity. One cannot add more species without eventually adding more functional groups as well. Another related determinant of productivity is the composition of the functional groups present: Productivity may depend not only on the number of such groups represented but also on their identities because some groups may be more important than others in contributing to productivity or resilience.

The average amount of biomass grown per year on a plot of a given size increased with the diversity of functional groups represented, as shown in Fig. 1. The increase leveled off after a certain point. Tilman and co-workers also found more nutrient uptake and better soil quality on plots with a more diverse collection of plant species. Furthermore, the plots that were more diverse in this sense were also more robust in the face of climatic fluctuations. It appears from this work that both functional diversity and species diversity are important in maintaining productivity and resilience. Having functionally similar plants that respond differently to environmental fluctuations contributes to resilience. It ensures that whatever the environmental conditions, there will be plants of a given functional type that thrive under those conditions. The functional composition of the community will therefore not be changed by environmental fluctuations. In contrast, if the members of different functional groups respond differently to environmental fluctuations, then these fluctuations will alter the functional composition of the community and therefore its ecological characteristics, and it will be less resilient in the face of such fluctuations (Chapin *et al.*, 1997). Chapin *et al.* (p. 503) argue that

> genetic and species diversity per se are important to long-term maintenance of community and ecosystem structure and processes. This argues that no two species are ecologically redundant, even if they appear similar in their ecosystem effects under one particular set of environmental conditions.

What is the mechanism behind these results? There is still dispute about this. One possibility is that each plant type has a range of climatic conditions to which it is best suited. Climate varies from year to year in terms of temperature, rainfall, and in many other ways, and if a plot contains only one plant type then in many years it will not have any plants well adapted to the climates of those years. If, however, a plot carries many types, then in most years there will be some that are well adapted to the climate of that year, and on average its productivity will be greater. This is another illustration of the old proverb, "Don't put all your eggs in one basket." Analytically, it illustrates the same point as the economic argument for holding a diversified portfolio of stocks. The more diversified the portfolio, the less vulnerable one is to conditions that are bad for particular stocks or stocks in a particular sector of the economy. A robust portfolio should have stocks that do well in times of growth, stocks that do well in times of high interest rates, and stocks that do relatively well in times of recession. It should have different types of stocks. Diverse plots will also be more resilient to climatic variation, as a part of the same phenomenon. This is related to the insurance role of biological diversity, here manifested as higher average productivity. An alternative explanation for the greater productivity of mixed plant communities has been suggested based on the interactions between the fungal communities associated with the roots of the plants and the diversity of the fungi in the soil, which affects the effectiveness of nutrient uptake (van der Heijden *et al.*, 1998; Read, 1998).

Similar studies have been conducted for microbial communities and have found similar results. Again, they show that more diverse communities are on average more stable and robust in the face of environmental fluctuations (McGrady-Steed *et al.*, 1997; Grime, 1977; Hooper and Vitousek, 1997; Naeem and Li, 1997).

There are other arguments about why diversity raises productivity, mostly specific to particular biological communities. Trees in forests provide a good example. Some trees are tolerant of shade and others of bright light. A forest of a uniform tree species will consist entirely of trees of one of these types. In contrast, consider a forest with tall shade-intolerant trees forming the upper canopy. These need bright light and in the upper canopy leaves receive this light. Below them are shorter trees that are more shade tolerant, and perhaps below these is another layer of even more shade-tolerant trees or shrubs. With such an arrangement of diverse species, bright light falls on those that most need it, and the light that passes through their canopy and is of lesser intensity then falls on plants well suited to it. Total photosynthesis—that is, conversion of carbon dioxide in the air to carbohydrates with the aid of light energy—will be greater under such a regime than in a single-species forest (Aber and Melillo, 1991).

These arguments show that diversity is important in ensuring the productivity and robustness of natural plant communities, and therefore of the ecosystems that are based on them. Diversity also helps natural ecosystems to make the best adjustments to conditions that vary over time or over space. Without the appropriate level of diversity, natural ecosystems cannot adjust to natural variations in the environment.

Through its role as the raw material in plant breeding, biodiversity also contributes substantially to the productivity of agricultural systems. New and higher yielding plant and animal varieties are generated from the natural variation in plants and animals. The large increases in grain yields of the "green revolution" of the 1960s and 1970s, which were responsible for keeping food output increasing in parallel with populations in developing countries, were largely achieved by use of genetic diversity in the plant populations. Estimates suggest that as much as $1 billion has been added to the value of the U.S. agricultural output each year for the past half century as a result of plant breeders' use of genetic diversity. Specifically, in the past half century there has been a doubling in yields of rice, barley, soybeans, wheat, cotton, and sugarcane; a threefold increase in tomato yields; and a quadrupling in yields of maize, sorghum, and potato (U.S. Congress Office of Technology Assessment, 1987). All this has been based on and derived from genetic variability in the underlying plant populations. In economic terms, this variability is an asset—and one that has yielded a high return at little cost. Diversity also contributes to productivity in agriculture through the practice of crop rotation. Changing from one crop to another from one year to the next can enhance soil productivity by increasing the nutrients in the soil. For example, rotating a crop such as wheat with a legume that fixes nitrogen can prevent long-term nitrogen loss and reduce the need for nitrogen fertilizers. This type of crop rotation was characteristic of traditional agricultural practices in medieval Europe.

It is important to note that although biodiversity contributes to the productivity of both natural ecosystems and agricultural systems, it does so through different mechanisms. Natural systems benefit directly from a diverse mix of species: Agricultural systems benefit from the existence of a pool of genetic variability on which breeders can draw. Agricultural systems are usually monocultures, consisting of a single species grown intensively over a large area. Its growth is supported by applications of water, fertilizers, pesticides, and weed killers. Farmers manage cropland so as to ensure that crop growth is not limited by lack of water or nutrients, and that the main food crop does not have to compete with other species or with pests during its growth. Farmers create and maintain an artificial environment and then plant a crop that is optimally adjusted to this environment—an approach that is radically different from the natural growth process.

II. BIODIVERSITY AND INSURANCE

A dramatic illustration of the insurance role of biodiversity comes from the recent history of rice production. The prosperity and comfort of literally billions of people depend on the rice harvest. In the 1970s, a new virus, the grassy stunt virus, carried by the brown plant hopper, threatened the Asian rice crop. This appeared to be capable of destroying a large fraction of the crop and in some years destroyed as much as one-fourth of the crop. Developing a form of rice resistant to this virus became of critical importance. Rice breeders succeeded in this task with the help of the International Rice Research Institute (IRRI) in the Philippines. The IRRI conducts research on rice production and holds a large seed bank of seeds of different varieties of rice and the near relatives of rice. In this case, the IRRI located a variety of wild rice that was not used commercially but which was resistant to the grassy stunt virus. The gene conveying resistance was transferred to commercial rice varieties, yielding commercial rice resistant to the threatening virus. This would not have been possible without genes from a variety of rice that was apparently of no commercial value. Without this variety, the world's rice crop, one of its most important food crops, would have been seriously damaged. Interestingly, the variety of wild rice that was resistant to the virus was found in only one location—a valley that was flooded by a hydroelectric dam shortly after the IRRI found and took into its collection the critical rice variety. The same situation was repeated later in the 1970s, and similar situations have occurred with other food crops, particularly corn in the United States (Myers, 1997). There is every reason to expect that events such as these will recur regularly: Planting large areas with genetically identical plants greatly increases the chances that once a disease starts it will spread with dangerous speed through the entire area and crop. A report by the Committee for Agriculture, Science and Technology (1999, p. 13) emphasized this point:

> Because of the increasingly high densities and large areas over which they are now grown, both livestock and crop plants are continually acquiring new diseases and pests, and existing diseases and pests are continually evolving new strains that overcome the defenses of particular breeds or strains. This is exacerbated by the accidental transport of diseases around the world. These diseases and pests destabilize agricultural systems. For instance, areas of western Minnesota and eastern North and South Dakota no longer can produce viable wheat and barley crops because of new strains of scab and vomit toxin for which no crop varieties have sufficient genetic resistance. Indeed, catastrophic attacks of disease, invasions of insects, and climatic extremes have caused wholesale crop destruction and ensuing famines whenever crops had insufficient diversity to provide at least some plants with the ability to withstand the assaults. Outbreaks of avian flu in the Chesapeake Bay area regularly result in rigorous quarantines of poultry houses due to the extremely high density of poultry farms in this region.
>
> Disease problems, as old as agriculture, are recorded in myth and in written history, and still exist. Red rust on wheat in Roman times, mass poisoning from ergot-tainted rye during the middle ages, the Irish potato famine of the 19th century, and the Southern corn leaf blight in 1970 all were due to insufficient biodiversity in the affected crops. The severity of the 1998 Hong Kong chicken epidemic was likely exacerbated by the lack of diversity in disease resistance as well as by the high chicken densities in the production facilities.
>
> The continual accrual of new diseases can be countered only if breeders can find sufficient genetic diversity within a crop or its relatives. Even the full complement of natural genetic variation, though, may not be sufficient to stop some diseases. Consider, for instance, the impacts of chestnut blight, an introduced disease that devastated what was once the dominant tree of the eastern United States, but which now occurs only as rare stump sprouts. Despite the vast geographic expanse and genetic diversity of the native North American chestnut, there is no known genetic resistance to its pathogen. Similarly, in vast areas of west and central Africa, livestock genetic resistance to the debilitating effects of trypanosomiasis is found only in a few unproductive local breeds. Despite massive efforts, the genetic mechanisms governing this resistance are not yet well understood.
>
> A lethal disease of corn, or wheat or rice, were it to appear, would devastate agriculture and human society. The only insurance that society has against such a catastrophe is biodiversity. Genetic diversity within a crop plant or animal species and its relatives might allow resistant strains to be discovered and used. Similarly, a diversity of potential food plants might allow another species to become an effective substitute for a major crop species that was lost to disease.

These cases illustrate clearly the insurance role of biodiversity. It is an important defense against disaster in the form of new diseases. The pathogens that cause disease are evolving continually, in an attempt to outwit our defenses against them. A clear example of this phenomenon is the evolution of antibiotic resistance among bacteria. The bacteria that cause several once common diseases in humans are now showing resistance to their principal controls, to the great concern of public health authorities. The same is happening with the pathogens that cause disease in crops and in commercial animals. Without reserves of genetic variability we may not be able to develop varieties of our agricultural crops and animals that can resist these new disease varieties. Indeed, it is precisely genetic variability in the pathogens that allows them to develop resistance. Genetic variability means that some of the disease-causing pathogens are naturally relatively unaffected by our defenses against them, which may be in the form of weed killers, insecticides, or vaccinations for livestock. These more resistant specimens are the ones that survive and from which new subsequent generations are bred. Therefore, pathogens use against us the mechanisms that we will use against them if we preserve and use genetic diversity. Without this diversity, we have disarmed unilaterally in the war against our most threatening enemies.

There is another important role for the insurance provided by biodiversity—to provide variability that could be critical in responding to the environmental changes wrought by humans. Human activity is changing the climate and the sea level, and it is making many more local changes in the environment. A hotter climate may require different crop varieties. An increase in sea level may lead to increased salinity in groundwater, and therefore to a need for crop varieties that are salt tolerant. A good example of the value of diversity in the context of a changed environment is the evolution of plants that have grown on mine wastes in the United Kingdom. These wastes are rich in heavy metals and are poisonous to most plant species (Antonovics *et al.*, 1971). The existence of a pool of genetic diversity allows us to find plants that could tolerate these poisons and even help to remove them from the soil. We are making changes to the global environment on an unprecedented scale, and biodiversity might be critical in allowing us to respond to the consequences of these changes. Population growth and environmental change mean that we now need the insurance provided by biodiversity more than ever before.

III. BIODIVERSITY AND GENETIC KNOWLEDGE

The third reason I previously gave for the importance of biodiversity is that it is a source of knowledge. We can learn from natural organisms how to make chemicals that have important and valuable properties. A good example is provided by the polymerase chain reaction (PCR). This reaction is central to culturing DNA specimens for analysis—as in forensic tests used in trials such as the O. J. Simpson trial, and in many processes central to the biotechnology industry. Culturing requires an enzyme that is resistant to high temperatures. Enzymes with the right degree of temperature resistance were found in hot springs in Yellowstone National Park, and the heat resistance of these was then used to create an enzyme that could be used to culture DNA specimens. This enzyme is now central to the rapidly growing biotechnology industry. There are many more less complex examples. In fact, 37% by value of the pharmaceuticals sold in the United States are or were originally derived from plants or other living organisms (Carte, 1996). Aspirin is derived from the bark of willow trees. The bark of yew trees has been used to derive a drug that is effective against ovarian cancer (Stierle *et al.*, 1993). A derivative of the rosy periwinkle flower is being used to cure childhood leukemia. The key point is that certain plants and animals are known to produce substances that are highly active pharmacologically. Plants that live in insect-infested areas produce substances that are poisonous to insects, and these have been used as the basis for insecticides. Some snakes produce venom that paralyzes parts of the nervous system, and others produce venom that reduces blood pressure. Other insects produce anticoagulants. All of these have been adapted for medical use. There is little that is new in these observations: They form the basis for many traditional medicines, which rely heavily on plants. Shakespeare refers to this in *Romeo and Juliet* (II: iii):

> O! mickle is the powerful grace that lies
> In herbs, plants, stones and their true qualities:
> For nought so vile that on earth doth live
> But to the earth some special good doth give,
> Within the infant rind of this weak flower
> Poison hath its residence and medicine power.

IV. BIODIVERSITY AND ECOSYSTEM SERVICES

I previously mentioned the role of genetic diversity in providing raw material for selective breeding, the traditional way of developing new crop or animal varieties that are more productive, more disease resistant, hardier, or more desirable in some other way. I have also mentioned its role in ensuring the productivity of ecosystems, and in ensuring their robustness against diseases and pathogens. There are other more complex ways in which biodiversity is essential to the proper functioning of ecosystems and to the delivery of the ecosystem services on which human beings are so dependent.

There are cases in which the full diversity of organisms in an ecosystem is required for that system to function and to provide services to human societies, and the removal or addition of even a single type of organism can have extraordinarily far-reaching consequences. "Keystone species" provide a convincing illustration. Ecologists use the term keystone species to describe a species whose removal will cause an entire ecosystem to change substantially. A widely cited example is that of sea otters on the California coast. The removal of sea otters as a result of hunting them for their pelts led to far-reaching and undesirable changes in the California coastal ecosystems. Sea otters eat sea urchins, which in turn graze on kelp plants. Without control of the urchin population by otters, the urchins will destroy the kelp beds, completely changing the marine coastal environment. Removal of otters led to a greatly impoverished coastal environment, which was restored in part to its original state with a ban on otter hunting. Another example of the role and impact of a keystone species is provided by the removal of kangaroo rats from an area of the Chihuahuan desert, which led to a threefold increase in the yields of grasses and to far-reaching changes in the desert ecosystem. In this case, the rats had played a key role by eating seeds and disturbing the soil, and their removal consequently changed the plant balance (Power *et al.*, 1996).

Not only can the removal of a species lead to major changes in an ecosystem but also the introduction of a new species (a so-called exotic species) can lead to a profound transformation of the system. A dramatic example is the introduction of the rinderpest virus into East Africa in 1890. This initially attacked domestic and wild cattle and then spread. By 1892, 95% of the wildebeest in the Serengeti region had died, in addition to most of the domestic cattle. Wildebeest are one of the main grazers and also the main food sources for carnivorous predators (lions, leopards, and hyenas) in the Serengeti; therefore, their virtual elimination led to profound changes in the system. In the 1930s, the introduction of a vaccination against rinderpest reestablished the original system (Aber and Melillo, 1991). The point of these examples is that we cannot easily tell a priori which species are essential and which are not. There is often a risk that an apparently small change in a set of species will have effects far beyond those initially anticipated. The degree of interdependence between different species is great; therefore, human beings may depend on many more species than we would expect from a first analysis of the situation. Abelard (an eleventh-century French theologian) suggested that any organism has a role to play and a reason for existing: "Whatever is generated is generated by some necessary cause, for nothing comes into being except there be some due cause for it" (as quoted in National Research Council, 1999). Also, John Donne (as quoted in Hemingway, 1940), an English metaphysical poet of the seventeenth century, wrote that

> No man is an island, entire of itself; every man is a piece of the continent, a part of the maine; if a clod be washed away by the sea, Europe is the less, as well as if a promontory were, as well as if a manor of thy friends or thine own were; Any man's death diminishes me, because I am involved in mankind; And therefore never send to know for whom the bell tolls: It tolls for thee.

There is an ecological equivalent to this: No species is an island, entire of itself, not even *Homo sapiens*. Any species' extinction may diminish us because we depend on many species. To repeat: The loss of even apparently unimportant species can have immensely costly consequences because of the complex patterns of interdependence between species. In the end, the loss of an apparently small and unimportant group of species could threaten the provision of ecosystem services that are essential to humanity. The distinguished biologist E. O. Wilson once said of microbes that "We need them but they don't need us." This is why many scientists see a serious risk in the current rate of species extinction: They cannot be precise about the dangers involved but nonetheless believe that there is a real risk of costly consequences. To give this point some substance, I mention a possible relationship between the extinction of passenger pigeons and the introduction of Lyme disease into American society. When Europeans first arrived in the United States, the passenger pigeon was probably the most abundant bird in the country. Its population was estimated in the billions. It traveled around in flocks of hundreds of thousands, flocks so large that their passing darkened the sky for many minutes at a time. By 1914 they were extinct, annihilated by a combination of hunting and destruction of the habitat that they needed for survival. It seemed unbelievable that an animal so abundant could be reduced to extinction so fast. A possible connection between this extinction and the emergence of Lyme disease events has recently been proposed. A letter to *Science* in 1998 made the following suggestion (Blockstein, 1998, p. 1831):[2]

> There is another possible twist to the complicated ecological chain of events presented by Clive G. Jones *et al.* (*Reports*, 13 Feb., p. 1023) whereby the incidence of Lyme disease might increase following population increases of mice allowed by a big mast year of acorns. A major competitor of deer and mice for these bumper crops has been absent from the eastern deciduous forests for a century. The extinct passenger pigeon (*Ectopistes migratorius*) was a nomadic wanderer that specialized on a diet of the superabundant, but unpredictable, crops of mast. With a population estimated at 2 to 5 billion, concentrated in enormous flocks, passenger pigeons congregated wherever there were huge crops of mast. The birds were so efficient at denuding the woods of nuts that many observers noted that native wildlife and feral hogs could not find sufficient food after a pigeon flock had passed through. Is it possible that, in the

---

[2] I am grateful to Paul Ehrlich for this reference.

> presence of passenger pigeons, the population explosions of mice in mast years, reported by Jones *et al.*, would have been less likely. Could the outbreaks of Lyme disease in the late 20th century have been a delayed consequence of the extinction of the passenger pigeon?

The point here is that passenger pigeons ate acorns and beechnuts, both of which were abundant in the forests of the northeastern United States. The demise of the pigeons led to an increase in the food available for other animals that ate these, including mice. Mice are the main breeding ground and hosts of the parasites that cause Lyme disease, and it is reasonable that the explosion of food for mice led to an increase in their population and thus in the population of Lyme disease vectors. (Abundant acorn crops always lead to increases in the population of mice.) The disease vectors transfer from mice to deer, which browse in the same forests and on the same foods and then move across territory likely to be used by humans, grazing on grass on lawns and fields. Therefore, the extinction of passenger pigeons could have been instrumental in causing the spread of Lyme disease to humans. This illustrates well the extraordinary complexity of the web of life and of the connections between different species and between species and human welfare. No one could reasonably have anticipated this connection between passenger pigeons and Lyme disease. No analysis of the consequences of the loss of this bird could have anticipated such an outcome. Indeed, the bird was so abundant that it must have been difficult to anticipate that human activity could drive it extinct.

The message of these examples is that it is difficult to foresee the consequences of a change in the biodivsity of an ecosystem. Even an apparently small change can lead to dramatic alterations in the system's ability to function and to provide the services on which human beings are dependent. There is another aspect of this phenomenon. A particular role in an ecosystem may be played at different times or in different circumstances by quite different plants or animals. The type of tree that stabilizes soil on a north-facing slope at a certain latitude may not grow on a south-facing slope at that latitude so that a different species is needed there to maintain the physical stability of the system. As a consequence, the set of species required for a certain type of ecosystem to function may vary greatly from region to region. In fact, we know of no single subset of species that on their own would serve to operate all ecosystems and provide all ecosystem services in all regions of the planet. Therefore, diversity in a given location may increase productivity and ecosystem functions in that location, whereas diversity at the regional or global level is actually necessary for the operation of important ecosystems in all geographic regions. Although individual species may possibly be redundant in some locations, it is possible that on the global scale few if any are really redundant. A clear statement on this topic is given by Chapin *et al.* (1997, p. 505):

> The abundance of species with similar ecological effects should give stability (resistance and resilience) to ecosystems in the face of increasingly rapid human-induced environmental change. Loss of a keystone species or of all species in a major functional group will, by definition, have large ecosystem effects. Efforts to identify and protect such species and groups often yield demonstrable near-term benefits. Of increasing concern is the loss of species that have similar ecosystem effects but differ in their environmental responses. Loss of such species may reduce ecosystem resilience and the capacity to adjust to ever-increasing rates of environmental change. This latter role of diversity is not adequately represented in current international conventions, but it may be one of the most important mechanisms by which we sustain the long-term functioning of ecosystems and the services they provide to society.

V. BIODIVERSITY AND MARKETS

Returning to the economic questions, to what extent can we hope to commercialize these contributions of biodiversity? Obviously, they are economically important, but can the market capture them? Can the economic contributions of biodiversity be used to generate incomes to the owners of biodiversity that will provide them with incentives to conserve it? Can they make conserving tropical forests more attractive than clearing them? This is a critical question: Forest owners do not conserve forest because they are important to humanity. Rather, if they conserve them, if they do so because they can profit from so doing. Therefore, we need to know whether the important services provided by biodiversity can be the basis for profits from forest conservation or for conservation of sources of biodiversity more generally. In this context, I again examine the categories under which I classified the contributions of biodiversity:

Increasing productivity
Providing insurance
Providing knowledge
Maintaining ecosystem services

One of the most fundamental insights into the operation of a competitive market economy is that under certain conditions it will align individual and social interests and provide incentives that lead to an efficient outcome. In the famous words of Adam Smith in his *An Enquiry into the Nature and Causes of the Wealth of Nations* (1977),

> Every individual ... neither intends to promote the public interest, nor knows how much he is promoting it. He intends only his own security, his own gain. And he is in this led by an invisible hand to promote an end which was no part of his intention. By pursuing his own interest he frequently promotes that of society more effectively than when he really intends to promote it.

This is a beautiful metaphor: Market forces are an invisible hand, steering us to act in the interests of society as a whole when in fact we only seek to promote our own interests. Recently, this insight was formalized and made more precise via some important propositions from economic theory, including the First Theorem of Welfare Economics. These propositions state that if all goods are private goods (i.e., there are no public goods) and the private and social costs of all activities are equal, then a competitive market economy is Pareto efficient (i.e., operates so as to leave unexploited no possibilities for mutual gain). This is a remarkable result and provides the basis for economists' belief in the efficacy of market systems and the desirability of market-based approaches to economic organization. However, in the context of biodiversity the restriction that all goods be private is critical.

What is the public–private good distinction? A public good is one that has two properties: My consumption does not interfere with yours, and the provider cannot prevent nonpayers from benefiting from the good. Such goods are said to be nonrival and nonexcludable, in contrast with private goods, which are both rival and excludable. A seat at the opera is a private good: If I sit in it you cannot, and the management can certainly exclude nonpayers from seeing the performance. Law and order, in contrast, is a public good: It benefits everyone in the region in which it is enforced and the benefits cannot be restricted to those who have contributed to its costs.

Markets are not good at providing public goods: Their nonexcludability makes it difficult for the provider to earn a good return on the costs of providing them. Knowledge is an interesting and relevant example: It is naturally a public good because it can be passed costlessly from one person to another and enjoyed by all of them, even though none of them, or perhaps only the first, paid for it. Hence the existence of intellectual property rights, instituted as a means of ensuring some return on the generation of knowledge. Because of the difficulty in appropriating the returns on their provision, markets tend to underprovide public goods relative to an economist's concept of what is efficient for society. Consequently, they have traditionally been provided by the public sector. Some of the services provided by biodiversity are public goods, although biodiversity does not fit the traditional mold for public goods completely. The extent of biodiversity is not something that can be determined by the public sector because it is the result of literally billions of land-use choices throughout the world. It is also strongly influenced by issues such as climate change, which are again driven by billions of heating and transportation choices. In fact, biodiversity has been called a privately produced public good. For the remainder of this argument, the key point is that the First Theorem fails in the context of public goods, and indeed it has long been recognized that markets will underprovide public goods relative to the level that would be required for efficiency.

One further basic economic point is needed—a familiarity with the diamond–water paradox and the limitations of market prices as indicators of "importance" to society. Discussion of this paradox helps us to be clear that the price of a good does not reflect its importance in any overall social or philosophical sense. Very unimportant goods can be valued more highly by the market than (have higher prices than) very important goods. The classic illustration of this is the diamonds and water paradox, which perplexed economists through the eighteenth and nineteenth centuries until its resolution by Alfred Marshall. The point here is that water is clearly more important to human society than diamonds, but diamonds trade in the market at prices far in excess of those fetched by water. Why? Marshall's answer was simple and is now part of common knowledge: Price is set by supply and demand. The market price is the price at which the amount supplied is also the amount demanded. In the case of water, the supply (at least in Marshall's time) was so large as to exceed the amount that could possibly be demanded at any price. Consequently, the price was zero: Water was free. Now, of course, the demand for water has increased greatly as a result of population growth and increasing prosperity, whereas the supply has remained approximately constant so that water is no longer free. For diamonds, being naturally scarce, the desire for ownership always exceeded that which could be accommodated naturally. The market price was high as a result of competition between rich people for the few diamonds available.

What are the implications for biodiversity? Simply that even if it is of great importance to society, and is not a public good, it will not necessarily be possible to convert this importance into value in the market place. The balance of supply and demand will be critical here, as shall be shown in the context of the commercialization of the genetic knowledge inherent in biodiversity.

A. Productivity

To the extent that diversity increases productivity in agricultural systems, we would expect that farmers would be willing to pay for it. Arable farmers achieve some of the benefits of diversity by crop rotation, i.e., rotating between a series of different crops in successive years. The different crop types make different demands on the soil and contribute different nutrients to it. However, the range of crops used for this purpose is quite limited and does not contribute in any substantial way to the conservation of biodiversity. The other mechanism through which diversity contributes to productivity in agriculture is via its contribution to the breeding of new plant species that are better adapted to emerging conditions or more resistant to new diseases. At this point, the productivity and knowledge roles of biodiversity merge, and I comment on the possibility of commercializing the knowledge role later. These comments on biodiversity's contribution to knowledge will also apply to the contribution that biodiversity makes to human societies via the breeding of new varieties.

There appears to be some appreciation of the benefits of diversity in tropical agriculture, in which there is a tendency to grow several crops together or to grow crops in a way that conserves the original forest. Traditionally, coffee was grown as an understory plant beneath high tropical forest trees: This benefits the coffee, which is a shade-tolerant plant, takes full advantage of the light available in the region, and avoids the need to destroy the main forest trees to make land available for coffee growing. This practice also allowed other commercial crops to be grown with the coffee, such as citrus fruits and avocados, thus allowing farmers to diversify their risks. In cases such as this there is a contribution to the preservation of diversity because of the conservation of the forest. Studies have shown that forests converted for production of shade-grown coffee retain a very large proportion of their original biodiversity and that growing in this way is less expensive per pound produced than plantation growing. This cost difference reflects in part the greater productivity of diverse ecosystems and the more effective cycling of nutrients in these plant communities (the total yield of coffee per hectare, however, is less; Perfecto *et al.*, 1996). In this case, it seems that there can be some conditions under which the productivity enhancements of diverse systems can be realized commercially, with attendant benefits for biodiversity conservation.

B. Insurance

Insurance is clearly something for which there is a demand. Most of us insure our homes and our cars and have health insurance. Therefore, perhaps the insurance role of biodiversity is one for which people will pay. The difficulty here is that until recently this insurance has been provided as a public good—indeed, as a global public good. Consider in the light of the public–private good distinction the conservation of rice varieties by the IRRI mentioned previously and in particular the use of one of these to provide a defense against the grassy stunt virus. In cases such as this, the insurance was provided as a public good. It was available to all for the cost of buying the new variety of rice incorporating resistance to the grassy stunt virus. You did not have to pay an insurance premium to benefit from the insurance. The developers of the new variety could not exclude from using it those who did not contribute to its development by paying insurance premiums.

At this point, I again digress briefly into economics. Risk, such as the risk of the destruction of a part of a crop, is managed in a market system by contingent contracts. A contingent contract is a contract that pays a specified amount if and only if a specified contingency occurs. Buying the contract reduces the risk to you that is associated with the specified event. If it occurs and has negative consequences, then you are to some degree compensated by the payment made to you under the contract. An insurance contract is a classic example of a contingent contract: You buy it by paying a premium and it entitles you to a payment if and only if a specified event—the insured peril—occurs. It is crucial for the efficient management of risks that the payment through which one purchases a contingent contract should not be contingent. You pay the insurance premium whether you use the insurance or not: The only uncertainty or contingency is whether you can make a claim under the policy.

It follows that the problem of the public good nature of the insurance provided by the existence of new varieties could not be overcome by charging premium prices for them. This price would only be paid after the product had been developed and needed. If the new variety were never needed, then this price premium would never be paid. No contingent contract is being sold in this framework.

There is a possibility that the public good nature of the insurance will change. Public goods can become private goods through either institutional or technical change. A good example is television broadcasting. Until about 10 years ago, this was a public good par excellence: A broadcaster could not exclude from viewing a program anyone in the reception area so that the nonexcludability property of public goods held, and of course there is no rivalry in consumption. My viewing a TV broadcast in no way interferes with your viewing. The development of scramblers changed this. A TV broadcast can now be scrambled so that it can only be viewed by those who have purchased a descrambler. Broadcasters can now exclude those who have not paid from viewing their programs, which are therefore no longer public goods. A public good has been privatized.

Developments in the area of intellectual property rights for agricultural biotechnology could change the situation in a similar way for the insurance value of biodiversity. These may lead to the privatization of a hitherto public good. Crop developers are increasingly patenting genes developed to enhance the properties of crops, including such properties as their taste, productivity, and insect resistance. Most agricultural biotechnology companies have a large and growing portfolio of patents on genes and on genetically modified plant varieties. They are also aggressively defending these patents, to the extent of developing and introducing "terminator genes" that will ensure that the properties conveyed by patented genes will not be transferred to offspring of the plant. Consequently, users cannot breed from seed that they have bought but must purchase more from the supplier. It is possible that plant breeders with such a tight hold on their intellectual property will be able to extract from users a great enough return to justify substantial investment in biodiversity conservation. The aggressive enforcement of intellectual property rights might act here like scramblers did with TV broadcasts, effectively privatizing a previously public good. Were it to happen, this would increase substantially the incentives for biodiversity conservation, but possibly at the cost of restricting access to the latest agricultural technology to those with the ability to pay a premium.

There is already a good illustration of the possible side effects of privatization of genetic knowledge. The bacteria *Baccilus thuringiensis (Bt)* produces a toxin that kills many crop pests and is itself biodegradable. For this reason, organic farmers use it as a pesticide: Because it is biodegradable it leaves no dangerous residues on the crops on which it is used. Monsanto and Novartis recently incorporated genes from *Bt* into transgenic crops. The presence of these genes in the plants means that the plants benefit from the defense provided by the *Bt* bacteria, which is the production of proteins that are toxic to the main pests of cotton and corn. Because the *Bt* genetic defense against pesticides is now widely used, it is possible that pests will develop resistance to it. Resistance develops faster the more widely a defense is used. Therefore, within a few years there may be generations of crop pests that are immune to this way of defending crops, i.e., immune to the *Bt* toxin. Monsanto and Novartis will seek to develop variants on the *Bt* genes, and with their scientific and financial resources they may succeed. If so, they will sell a new generation of transgenic plants with defenses against the new generation of pests. However, these will be their proprietary products, covered by their patents. Because of the development of resistance to the *Bt* toxin, organic farmers will no longer be able to use *Bt* as a harmless pesticide. They will be forced to buy proprietary defenses against pests. This is a clear illustration of the two factors mentioned previously: development of intellectual property rights in genetic knowledge leading to both stronger incentives to conserve and to develop further and also leading to a restriction of access. In the case of transgenic crops including the *Bt* defenses, Monsanto recently announced that all growers of *Bt* corn will be required to grow plots of nonengineered corn that are at least 20% of the size of the engineered crops. The aim here is to provide sufficient nonengineered corn that the development of resistant pests will be delayed or possibly even prevented. ("Monsanto Concession," 1999).

C. Knowledge

Here, I discuss the role of biodiversity as a source of knowledge. There are two different areas of application here: the development of medical products by the pharmaceutical industry and the development of new or better crops by the agricultural biotechnology industry.

In applications of biodiversity to the development of pharmaceuticals there has already been some progress toward commercialization. Recognition of the likelihood that tropical plants contain chemicals that could be forerunners of pharmaceuticals has led most major drug companies to pursue bioprospecting as a way of finding new pharmacologically active substances to serve as a basis for drug development. Typically, they have sought these compounds in the tropics, in areas where there is extensive interspecies competition, or in other extreme areas. They have been willing to pay quite substantial sums for access to these regions, and they have made deals with host countries that involve giving them a royalty on the products that might eventually be based on their prospecting. Such royalties could be large relative to the incomes of the countries concerned. Merck, Inc., one of the largest pharmaceutical companies in the United States, has an agreement with a Costa Rican agency called InBio (Institute Nacional de La Biodiversidad) for bioprospecting rights in Costa Rica. The terms of the agreement are that Merck paid InBio a fixed sum, $1.35 million, to be used for forest conservation in exchange for the right to receive samples collected by InBio and to use these as the basis for new product development. Should any of them prove commercially successful, Merck will pay InBio a royalty on the revenues generated. Similar agreements are in place between other U.S. pharmaceutical companies and regions of Central and South America.

The discovery of the PCR enzyme and the agreement between Merck and InBio and several other drug discoveries based on plants from developing countries led to a wave of optimism, perhaps excessive, about the potential commercial value of *in situ* biodiversity in developing countries. What in fact is the commercial potential here? There is no question that pharmaceutical and agricultural products of great human and commercial value have been and will in the future be developed from the biodiversity in tropical countries. The key question is how much of this will be returned as a reward for the conservation of the originating biodiversity. In answering this question, one must take note of several points. The first point is related to the discussion of the diamond–water paradox: Not everything that is important will have value in the marketplace because there is nothing associated with being important that rules out the possibility that the supply may exceed the demand. Some commentators have seen this as a critical issue in the market for biodiversity-based genetic information. There are literally millions of organisms in the world that might provide genetic information, and if we do not know which ones will provide valuable information and which will not, then the supply of potential genetic leads is huge, possibly greater than pharmaceutical companies can process. In such a situation, the market price for such leads would be near zero. However, recent calculations suggest that in a small number of the world's biodiversity hot spots, bioprospecting rights may be worth as much as $9000 per hectare. This is small in relation to the amount that might ultimately be derived from drug sales but still large relative to other uses of the land. In fact, it is approximately a century's worth of ranching income (Rausser and Small, 1998). The key insight in these calculations is that prior knowledge of the nature of the ecosystems in a location can improve estimates of the probability of finding commercially interesting compounds there and can suggest where search would be profitable and where success is unlikely. In this case, not all organisms or all leads are equivalent: Some have much greater chances of success than others. Prior scientific information can change the odds of success from one 1 in 10,000 to approximately an order of magnitude better. In practical terms, this means that developing countries can clarify the commercial attractions of their biodiversity by performing research on the ecosystems of which it is a part. This is similar to a country with potential oil reserves engaging in basic geological prospecting before seeking to negotiate leases for oil development. The results may be positive or negative, but either way they will give the country a better view of its prospects. In those cases in which the research is positive, the impact on the value of prospecting rights could be large.

Another important point is that an immense amount of human skill and expertise are needed to develop a plant specimen into a commercial drug. Typically, there will be a minimum of 10 or more years of work by hundreds of skilled people working with millions of dollars of sophisticated equipment. Unfortunately, very few plant extracts actually produce drug leads (i.e., contain pharmacologically active compounds with no obvious ill effects)—probably less than 1 in 10,000. Of these few, very few become commercial drugs—less than 1 in 100. On average, perhaps 1 in 250,000 samples collected leads to a commercial drug. ("When Rhetoric Hits Reality," 1998).

Therefore, the chance of any individual bioprospecting operation leading to a commercially valuable drug is very small indeed. Also, even if it does produce a drug, tens or even hundreds of millions of dollars will have to be invested, with significant chances of failure. Developments in biotechnology are currently altering this picture. They are reducing the time needed for testing and development and giving greater insights into the kinds of chemicals likely to be successful. By reducing the costs of drug development based on bioprospecting, they are making bioprospecting more attractive. For example, the cost of screening 10,000 samples for pharmaceutical potential 10 years ago would have been $6 million: Today it is $150,000 (Reid *et al.*, 1993). Simultaneously, advances in knowledge are also making more effective alternative methods of drug development, based on understanding of the cellular and genetic mechanisms of disease. In total, the picture that emerges is one of heavily guarded optimism. Bioprospecting does have economic value, and technological developments may be increasing that value. However, in the short term we cannot expect great sums of money to flow to the conservation of biological diversity because of bioprospecting possibilities.

There is another problem to be overcome in establishing an income flow from bioprospecting—a problem with intellectual property rights. The same plant may occur in several different regions, and the same or similar chemicals may occur in different plants. Therefore, the same or similar drugs may be derived by different routes from different plants or different geographic regions. Research toward a commercial product has to be well under way before it is patentable; therefore, there is always a risk of being blocked by a prior patent.

There is also a risk, regarding the conservation of biodiversity, that biodiversity is valuable but leads to no direct commercial application. In a recent article on the value of marine bioprospecting, Carte (1996, p. 284) stated,

> Although many of these products are not likely to become therapeutics, the information gained from studying them is likely to lead to the development and understanding of novel molecular targets, which may in turn lead to the development of new therapeutic agents.

This is a classic statement of the importance of basic knowledge. Basic knowledge is a public good: Its importance is great but it is not patentable and not something that can be appropriated by a typical bioprospecting contract with an element of royalty payment or revenue sharing. The development of basic knowledge is typically publicly funded precisely because its economic benefits, although potentially immense, are difficult to appropriate. More constructive thinking about how to realize the undoubted importance of bioprospecting in terms of income for conservation is needed. A further illustration of this point is provided by the example of taxol mentioned earlier. Taxol is a promising antitumor agent in breast and ovarian cancers that can be extracted from a fungus that lives in the phloem (inner bark) of the Pacific yew tree. Taxol was first isolated from the tree itself, but the tree is relatively rare and slow growing, and it produces little taxol; therefore, a search for other sources was initiated (Stierle *et al.*, 1993). Ultimately, little in the way of economic returns may flow to the regions in which taxol was discovered.

Biodiversity can also be applied to the development of new or better crops: To date, this has probably been the most important commercial application of biodiversity. As noted previously, the existence of a pool of genetic variation provides plant and animal breeders with the raw materials for developing new varieties and more productive or resilient variants of existing varieties. The existing varieties of a commercially important crop are usually the property of a commercial firm or of a research facility and are protected by patents. For example, the University of California owns the patents of many varieties of strawberries. These patents cover varieties that are best suited to different soil types and different weather conditions, that are least prone to spoiling during transport and storage, and so on. For most other commercial crops, the varieties are owned by seed companies, whose main asset is often the intellectual property represented by their ownership of patents to widely used varieties. In this context, the market can certainly recognize the value of biodiversity, provided that it is of the type that seems likely to contribute to the development and refinement of commercially important crops. Unusual variants of commercial crops, such as early variants of wheat, corn, or soybeans, would qualify, and possibly so too would their near genetic relatives. However, biodiversity more broadly would probably not derive a value through this process, even though genes from unrelated plants might enhance the commercial potential of existing crops.

D. Ecosystem Services

Perhaps the most promising approach is to consider selling the services of natural ecosystems and using the revenue to provide incentives for conserving the biodiversity that supports them. Selling services provided by natural ecosystems can potentially provide incentives for the conservation of these systems and thus indirectly of biodiversity. For example, watersheds provide economically valuable services for which there is a market, and indeed recognition of this has already led to the conservation of significant forest areas (Chichilnisky and Heal, 1998). Some commentators (Reid, 1998) have suggested that as much as 15% of the earth's land area serves as watersheds for large cities and so could legitimately be conserved on the basis of the watershed services that it supplies. Many of these watersheds are areas of considerable biological uniqueness and their conservation would be a major advance for the conservation of biodiversity. Their functioning as watersheds probably depends substantially on the continuation of their current levels of biodiversity.

Ecotourism is based on the preservation of intact ecosystems and the more appealing elements of biodiversity to be found in some of these systems, and revenues from this are providing powerful incentives for the conservation of several important ecosystems (Chichilnisky and Heal, 1998; Heal, 2000a). Insofar as ecotourism conserves tropical ecosystems, it contributes to the conservation of biodiversity. In certain regions, particularly southern Africa, ecotourism is making a major contribution to the conservation of biodiversity. In Angola, Botswana, Kenya, Malawi, Mozambique, Namibia, South Africa, Tanzania, Zambia, and Zimbabwe, approximately 18% of the total land area is now devoted to the support of wildlife. This is both a significant amount in total and a massive increase relative to 15 or 20 years ago (Bond, 1993; Cumming, 1990a, and b; Cumming and Bond, 1991). This has led to an increase in the populations of several previously endangered species and to more robust breeding populations of many important birds and mammals. All of this has been driven by the fact that in many parts of this region land devoted to ecotourism and sport hunting can earn a higher return than land devoted to more conventional agriculture. Figure 2 illustrates these phenomena: It shows for Zimbabwe both the growth increase in land whose primary use is the support of wildlife and the favorable economic returns to wildlife conservation relative to agriculture in parts of the region. (Regions III–V are either arid regions or wetlands.) Although the growth of ecotourism as an economically important activity has increased most in southern Africa, this region is not unique in being able to generate conservation incentives by tourism. Central and South America, and parts of Asia, are now benefiting from ecotourism, which is providing significant returns to the conservation of certain endangered species in these regions (Heal, 2000b; Freese, 1999; Honey, 1999).

Finally, there is a real prospect of commercializing some of the carbon sequestration services of forests under the terms of the Kyoto Protocol, which provides for compensation for some carbon sequestration activities via its provisions for joint implementation and through its clean development mechanism. I have estimated elsewhere (Heal, 2000b) that this could generate

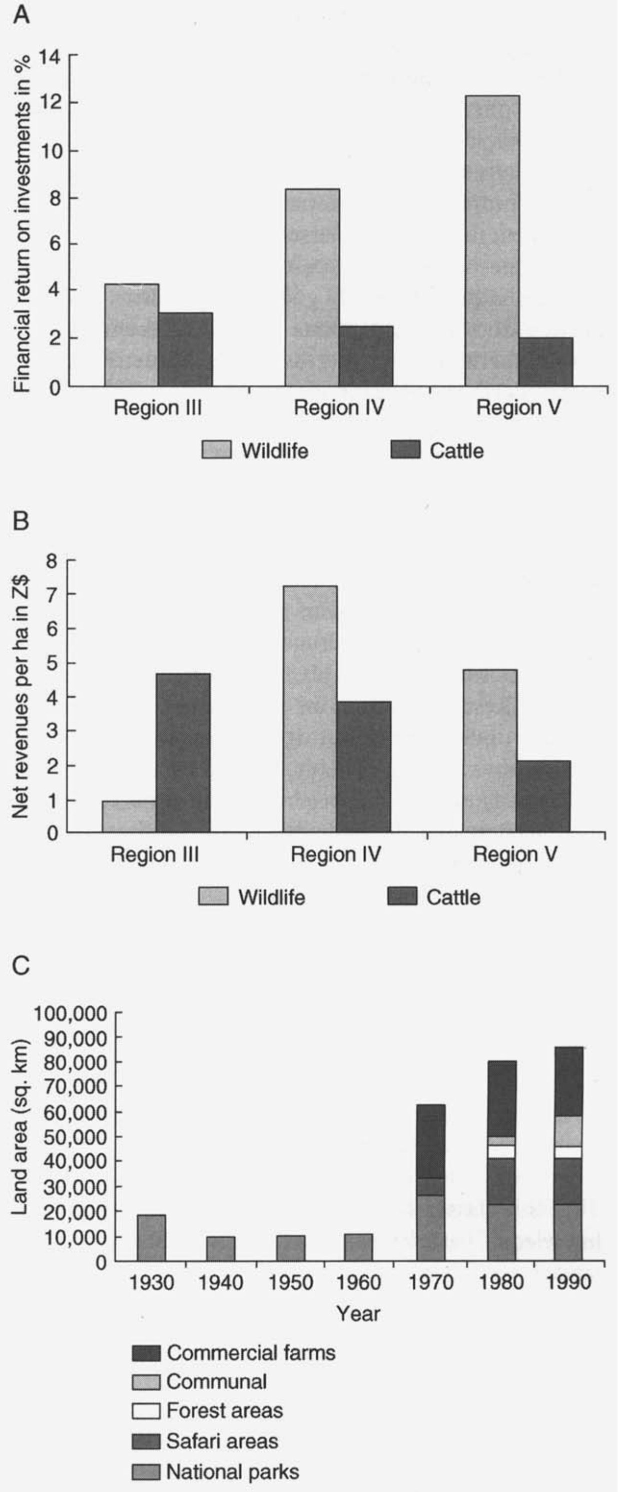


FIGURE 2 (A) Financial returns to wildlife and cattle ranching in Zimbabwe. (B) Returns per hectare from cattle and wildlife in Zimbabwe. (C) Land area (square kilometers) for wildlife conservation in Zimbabwe.

an income as high as $50–100 per hectare per year, which is high enough to radically change the incentives for forest conservation and hence biodiversity conservation.

There are also other economic mechanisms through which the goods and services provided by tropical forests and their associated diversity might be marketed, although it is not clear whether these mechanisms can be applied generally. One is the sale of "nontimber forest products," which are commercially valuable forest products that are not produced by tree cutting. They are therefore not timber, and their harvest and sale are compatible with conservation of the forest. They include various tropical fruits, vines that can be used as ropes, rattan (which grows among the trees), resins such as latex, and plants used as medicine by local populations. This last use is important. Approximately 4 billion people have no access to Western-style medicine and in times of sickness depend on plant extracts; this is known as ayurvedic medicine. This market is important in human terms, and although the amounts of money involved are small by comparison with those in Western medicine, they could still be enough to provide a significant incentive. Recent estimates suggest that in some tropical forests the collection and sale of nontimber forest products could yield as much as $60–140 per hectare per year, although many other studies have suggested significantly lower numbers (Grimes *et al.*, 1994; Houghton and Mendelsohn, 1996; Peters *et al.*, 1989). These are significant amounts. Given that they are sustainable, i.e., can be earned year after year, they are probably sufficient to justify conservation of forests even in the face of pressure for logging. However, currently there are a limited number of studies on this form of forest use so that the generalizability of these numbers is not clear, and whether the implied levels of harvesting are truly sustainable is also unclear.

The different mechanisms for generating income from natural ecosystems and the biodiversity that powers them are not mutually exclusive: The same land area could earn income by all mechanisms. A forest could obtain returns from carbon sequestration, bioprospecting, nontimber products, managing a watershed, and ecotourism. In fact, the region of the Mata Atlantica (Brazilian coastal rain forest) inland from Rio de Janeiro is in a position to do exactly this. It manages the watershed for Rio in much the way that the Catskills region does for New York. It also manages the stream flow of the river Rio Paraibo do Sul, which provides most of the electric power for Rio via hydropower. These two services make it truly a major utility for Rio, with great economic value. Additionally, it supports a wide range of endemic species, sequesters carbon, and acts as a magnet for tourists. Currently, the region obtains a financial return only on one of these activities—ecotourism. In a case such as this, it is clear that the economic incentives for conservation could be immense, if we were only to do effectively what we already know how to do.[3]

VI. CONCLUSIONS

What can we conclude about biodiversity, its economic importance, and the prospects for generating income based on this? Clearly, its importance is great, in the sense that it has been and remains a key contributor to human well-being. We understand enough of this contribution for its magnitude to be clear, although there are probably many aspects that are still unknown to us. Economically, the following is a key question: Can this importance be reflected in a commensurate income yielded by the conservation and use of biodiversity and of goods and services based on it?

In essence, much of what biodiversity contributes is, or has traditionally been, a public good. Its contribution to the functioning of natural ecosystems is nonexcludable and nonrival, as is its insurance value. Also, knowledge is the quintessential public good. Despite the overwhelming publicness of much that biodiversity offers, there is some prospect of commercializing a limited part. Some of the services provided by natural ecosystems can be privatized and sold, generating a return to the conservation of the biodiversity that supports them. In the case of knowledge, the prospect of commercialization rests on the ability to establish intellectual property rights that will effectively privatize some of the public good aspects of its insurance and knowledge functions. Probably the most fundamental of these functions, like basic scientific research and development, will never be privatized and will thrive only with financial support via other mechanisms.

In summary, biodiversity is important economically. There are some market-based approaches to obtaining returns from biodiversity that can provide incentives for its conservation. If fully implemented, they would

---

[3] All of these measures are consistent with the United Nations Convention on Biological Diversity, to which many nations (although not the United States) have subscribed. This states in article 11 that "Each Contracting Party shall, as far as possible and appropriate, adopt economically sound measures that act as incentives for the conservation and sustainable use of biological diversity."

have a profound and positive impact. However, they would not solve the problem of biodiversity conservation. They would need to be supplemented by nonmarket measures such as the Endangered Species Act (ESA), the Convention on International Trade in Endangered Species (CITES), the Marine Mammals Protection Act, the Fishery Conservation and Management Act, and others. I am not suggesting that these are good examples of nonmarket policies: Indeed, in several cases they are not. The point is that they illustrate the type of option available as an alternative to the market. Of course, except for CITES, the previously listed measures are all domestic U.S. measures. To be really effective, we would need global equivalents—a Global Endangered Species Act, for example. Given the political complexity of the domestic ESA, this prospect should illustrate clearly the desirability of market-based policies whenever possible.

It is important to note that even regulatory approaches such as the ESA can give rise to market-based incentives if coupled with provisions for measures such as mitigation banking. In the case of the ESA, several states have modified this in a way that has proven highly effective and has introduced economic incentives where none existed in the initial formulation of the act. A good illustration is an agreement reached between the International Paper Corporation and the U.S. Fish and Wildlife Service concerning the red cockaded woodpecker. This bird is endangered and nests in forests owned by International Paper. The U.S. Fish and Wildlife Service and International Paper reached the following arrangement: A target number of breeding woodpecker pairs was agreed on, and provided that this number is attained or exceeded, then International Paper will be regarded as complying with the ESA, whatever modifications it might make in the habitat under its control. Furthermore, the agreement also provided that any surplus of breeding pairs over this number could be "banked." This means that it could be used by the company to offset ESA requirements with respect to red cockaded woodpeckers elsewhere, or title to the surplus could be sold to other landowners and used by them to gain some measure of exemption (Jorling, 2000; Noss *et al.*, 1997). The important point is that the costs of compliance with the ESA have been reduced by this agreement, without reducing its effectiveness. Indeed, there are additional benefits: Because the production of nesting pairs over a target level is saleable, International Paper now actually has an economic incentive to encourage the endangered species, something it never had with a strict interpretation of the ESA. Recent reports suggest that International Paper may be able to sell banked breeding pairs for as much as $100,000 per pair (W. Coleman, personal communication). Similar mitigation banking systems have been put in place for wetlands, which have to be conserved under the Clean Water Act and which also provide important habitat for endangered or threatened species.

In the cases of the red cockaded woodpecker and wetlands, regulation has produced a market. The market occurs as a result of efforts to meet the regulations at minimum costs and has the positive effect of providing stronger incentives than the original regulation for the conservation of the endangered species or habitat. In principle, we do not need to go through the process of regulation to set up a market. The most obvious move from an economic perspective is instead to create a market in situations in which this seems needed by using the state as a buyer. In other words, in the case of the red cockaded woodpecker, the government could just pay landowners on whose land the woodpeckers breed. This immediately establishes an incentive for conservation without the political and other costs of regulation. Systems of this type are currently being used on a trial basis in parts of the United Kingdom. In regions in which hill farming is economically marginal, and has been subsidized by the European Community's Common Agricultural Policy, the farming subsidies have been replaced by subsidies for preserving the ecological integrity of the countryside.[4] Specifically, farmers are paid to grow and maintain hedgerows, ponds, wetlands, wildflowers, and coppices that provide habitat for birds and mammals. The *Financial Times*, the European equivalent of the *Wall Street Journal*, recently commented in an editorial that in many cases it would make more economic sense to pay farmers to provide "ecological sustenance" rather than to grow food that is clearly surplus to the region's needs and cannot compete on world markets ("Too Much Food," 1998). This solution is less attractive than establishing a market in which consumers buy directly, as in the cases of ecotourism or watersheds, because the money spent by the government has to be raised as tax revenue from individuals or corporations. This can produce a loss of economic efficiency elsewhere in the economy. However, if the services provided by biodiversity are public goods, then there may be little alternative to having the public sector act as a buyer on behalf of society as a whole.

---

[4] This system, known as Tyr Cwmen, is in operation in parts of South Wales.

See Also the Following Articles

DEFORESTATION AND LAND CLEARING • ECONOMIC GROWTH AND THE ENVIRONMENT • ECONOMIC VALUE OF BIODIVERSITY, OVERVIEW • ECOSYSTEM SERVICES, CONCEPT OF • KEYSTONE SPECIES • LAND-USE ISSUES • MARKET ECONOMY AND BIODIVERSITY • PLANT SOURCES OF DRUGS AND CHEMICALS

Bibliography

Aber, J. J. D., and Melillo, J. M. (1991). *Terrestrial Ecosystems.* Saunders, Philadelphia.

Antonovics, J., Bradshaw, A. D., and Turner, R. G. (1971). Heavy metal tolerance in plants. *Advances in Ecological Research* 7, 1–85.

Blockstein, D. (1988). Lyme disease and the passenger pigeon. *Science* 279, 1831.

Bond, I. (1993, April). The economics of wildlife and land use in Zimbabwe: An examination of current knowledge and issues. World Wildlife Fund (WWF) Multispecies Animal Production Systems Project, Project Paper No. 36. WWF Program Office, P.O. Box CY 1409, Causeway, Harare, Zimbabwe.

Carte, B. K. (1996). The biomedical potential of marine natural products. *BioScience* 46(4), 271–286.

Chapin, F. S., III, Walker, B. H., Hobbs, R. J., Hooper, D. U., Lawton, J. H., Sala, O. E., and Tilman, D. (1997). Biotic control over the functioning of ecosystems. *Science* 277, 500–504.

Chichilnisky, G., and Heal, G. M. (1998). Economic returns from the biosphere. *Nature* 391, 629–630.

Committee on Agriculture Science and Technology (CAST) (1999, February). Benefits of biodiversity, Task Force Report No. 133. CAST, 4420 West Way, Ames, IA 50015-3447.

Cumming, D. H. M. (1990a, June). Developments in game ranching and wildlife utilization in East and Southern Africa. World Wildlife Fund (WWF) Multispecies Animal Production Systems Project, Project Paper No. 13. WWF Program Office, Zimbabwe.

Cumming, D. H. M. (1990b, June). Wildlife and the market place: A view from Southern Africa. World Wildlife Fund (WWF) Multispecies Animal Production Systems Project, Project Paper No. 12. WWF Program Office, Zimbabwe.

Cumming, D. H. M., and Bond, I. (1991, July). Animal production in Southern Africa: Present practices and opportunities for peasant farmers in arid lands. World Wildlife Fund (WWF) Multispecies Animal Production Systems Project, Project Paper No. 22. WWF Program Office, Zimbabwe.

Daily, G. C. (1997). *Nature's Services: Societal Dependence on Natural Ecosystems.* Island Press, Washington, D.C.

Ehrlich, P. R. (1988). The loss of diversity: Causes and consequences. In *Biodiversity* (E. O. Wilson, Ed.). National Academy of Sciences/Smithsonian Institution, Washington, D.C.

Freese, C. H. (1999). *Wild Species as Commodities: Managing Markets and Ecosystems for Sustainability.* Island Press, Washington, D.C.

Grime, J. P. (1977). Biodiversity and ecosystem function: The debate deepens. *Science* 277, 1260.

Grimes, A., Loomis, S., Jahnige, P., Burnham, M., Onthank, K., Alarcon, R., Cuenca, W. P., Martinez, C. C., Neill, D., Balick, M., Bennet, B., and Mendelsohn, R. (1994). Valuing the rain forest: The economic value of nontimber forest products in Ecuador. *Ambio* 23(7), 405–410.

Heal, G. M. (2000a). Sustainability and markets. In *Topics in Environmental Law* (R. Stewart, Ed.). Cambridge Univ. Press, Cambridge, UK.

Heal, G. M (2000b). *Earthkeeping.* Island Press, Washington, D.C.

Hemingway, E. (1940). *For Whom the Bell Tolls.* Scribner, New York.

Honey, M. (1999). *Ecotourism and Sustainable Development: Who Owns Paradise?* Island Press, Washington, D.C.

Hooper, D. U., and Vitousek, P. M. (1997). The effect of plant composition and diversity on ecosystem processes. *Science* 277, 1302–1305.

Houghton, K. T., and Mendelsohn, R. O. (1996). An economic analysis of multiple-use forestry in Nepal. *Ambio* 25(3), 156–159.

Jorling, T. C. (2000). Incentive-based management strategies: A "common-sense" approach to conservation of biological diversity. *Sci. Total Environ.*, 240, 1999.

McGrady-Steed, J., Harris, P. M., and Morin, P. J. (1997). Biodiversity regulates ecosystem predictability. *Nature* 390, 162–165.

Monsanto concession on engineered corn (1999). *Nature* 397, 98.

Myers, N. (1997). Biodiversity's genetic library. In *Nature's Services: Societal Dependence on Natural Ecosystems* (G. C. Daily, Ed.). Island Press, Washington, D.C.

Naeem, S., and Li, S. (1997). Biodiversity enhances ecosystem reliability. *Nature* 390, 505–509.

National Research Council (NRC) (1999). A framework for managing biodiversity. Report of the Committee on the Economic and Non-Economic Value of Biodiversity. NRC, Washington, D.C.

Noss, R. F., O'Connell, M. A., and Murphy, D. D. (1997). *The Science of Conservation Planning: Habitat Conservation under the Endangered Species Act.* Island Press, Washington, D.C.

Perfecto, I., Rice, R. A., Greenberg, R., and van der Voort, M. E. (1996). Shade coffee: A disappearing refuge for biodiversity. *BioScience* 46(8), 598–608.

Peters, C. M., Gentry, A. H., and Mendelsohn, R. O. (1989). Valuation of and Amazonian rainforest. *Nature* 339, 655–656.

Power, M. E., Tilman, D., Estes, J. A., Menge, B. A., Bond, W. J., Mills, L. S., Daily, G. C., Castilla, J. C., Lubchenco, J., and Paine, R. T. (1996). Challenges in the quest for keystones. *BioScience* 46(8), 609–620.

Rausser, G. C., and Small, A. A. (1998, March). Valuing research leads: Bioprospecting and the conservation of genetic resources. Paper presented at the Conference on Managing Human-Dominated Ecosystems at the Missouri Botanical Gardens. St. Louis, MO.

Raven, P. H., and Williams, T. (1999). *Nature and Human Society: The Quest for a Sustainable World.* National Academy of Sciences, Washington, D.C.

Read, D. (1998). Biodiversity: Plants on the web. *Nature* 396 (News and Views).

Reid, W. V. (1998, March). A business plan for ecosystem services: Extending the New York City watershed model to other geographic regions and other ecosystem services. Paper presented by the World Resources Institute at the Conference on Managing Human-Dominated Ecosystems at the Missouri Botanical Gardens. St. Louis, MO.

Reid, W. V., Laird, S. A., Meyer, C. A., Gamez, R., Sittenfeld, A., Janzen, D. H., Gollin, M. A., and Juma, C. (1993). *Biodiversity Prospecting: Using Genetic Resources for Sustainable Development.* World Resources Institute, Washington, D.C.

Smith, A. (1977). *An Enquiry into the Nature and Causes of the Wealth of Nations.* Chicago: Univ. of Chicago Press, Chicago.

Stierle *et al.* (1993). Taxol and taxane production by *Taxomyces* *andreanae*, and endophytic fungus of Pacific Yew. *Science* 260, 214–216.

Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., and Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science* 277, 1300–1305.

Tilman, G. D., and Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature* 367, 363–365.

Tilman, G. D., Wedin, D., and Knops, J. (1996). Productivity and sustainability influenced by biodiversity in grassland ecosystems. *Nature* 379.

Too much food (1998, November 17). *Financial Times*.

U. S. Congress Office of Technology Assessment (1987). U. S. Congress, Washington, D.C.

Van der Heijden, M. G. A., Klironomos, J. N., Ursic, M., Moutoglis, P., Streitwolf-Engel, R., Boller, T., Weimken, A., and Sanders, I. R. (1998). Mycorrhizal fungal diversity determines plant biodiversity, ecosystem variability and productivity. *Nature* 396, 69–71.

When rhetoric hits reality in debate on bioprospecting (1998). *Nature* 392, 535–540.

Wilson, E. O. (Ed.). (1988). *Biodiversity*. National Academy of Sciences/Smithsonian Institution, Washington, D.C.

BIODIVERSITY, DEFINITION OF

Ian R. Swingland
The Durrell Institute of Conservation and Ecology

GLOSSARY

biodiversity/biological diversity Species, genetic, and ecosystem diversity in an area, sometimes including associated abiotic components such as landscape features, drainage systems, and climate.

diversity indices Measures that describe the different components of biodiversity, such as species richness (alpha diversity), beta and gamma diversity, endemicity, and higher taxon richness.

ecosystem diversity Diversity of habitats, ecosystems, and the accompanying ecological processes that maintain them.

endemicity State of a species or other taxon being restricted to a given area, such as a specific habitat, region, or continent.

flagship species Charismatic or well-known species that is associated with a given habitat or ecosystem and that may increase awareness of the need for conservation action.

genetic diversity Genetic variety found within or among species; this diversity allows the population or species to adapt and evolve in response to changing environments and natural selection pressures.

keystone species Species that has a disproportionately greater effect on the ecological processes of an ecosystem, and whose loss would result in significantly greater consequences for other species and biotic interactions.

organismal (species) diversity Number and relative abundance of all species living in a given area.

species richness Absolute number of species living in a given area (also called alpha diversity), giving equal weight to all resident species.

use values Values that are obtained by using a natural resource, such as timber, fuelwood, water, and landscapes. These include direct, indirect, option, and nonuse values.

THE WORD BIODIVERSITY IS A MODERN CONTRACTION OF THE TERM BIOLOGICAL DIVERSITY. Diversity refers to the range of variation or variety or differences among some set of attributes; *biological* diversity thus refers to variety within the living world or among and between living organisms.

I. WHAT IS BIODIVERSITY?

The term “biodiversity” was first used in its long version (biological diversity) by Lovejoy (1980) and is most commonly used to describe the number of species. Recognizing that conventional methods of determining, and separating, species were inadequate, others elaborated the definition by including the variety and variability of living organisms.

These reduced and simple definitions, which embrace many different parameters, have been much elaborated and debated in the last three decades (see Section II); upon this definition hangs the outcome of important scientific considerations, international agreements, conventions, conservation initiatives, political debates, and socio-economic issues. Indeed, while the word “biodiversity” has become synonymous with life on earth, the term is commonly used in the fields of politics and environmental technology in addition to various scientific disciplines (Ghilarov, 1996). The U.S. Strategy Conference on Biological Diversity (1981) and the National Forum on Biodiversity (1986) in Washington, D.C., were the critical debates in crafting a definition, and it was the proceedings from the latter, edited by E. O. Wilson, that “launched the word ‘biodiversity’ into general use” (Harper and Hawksworth, 1994).

In measuring biodiversity, it is necessary to deconstruct some of the separate elements of which biodiversity is composed. It has become widespread practice to define biodiversity in terms of *genes, species, and ecosystems,* for example, “the abundance, variety, and genetic constitution of native animals and plants” (Dodson *et al.,* 1998). Biodiversity also encompasses all five living kingdoms, including fungi. However, biodiversity does not have a universally agreed on definition and it is often re-defined on each occasion according to the context and purpose of the author.

II. DEFINITION OF BIODIVERSITY

“Biodiversity” is a relatively new compound word, but biological diversity (when referring to the number of species) is not. Over the last decade its definition has taken a more reductionist turn. Possibly the simplest definition for biodiversity, lacking in specificity or context, is merely the number of species. Yet many have argued that biodiversity does not equate to the number of species in an area. The term for this measure is species richness (Fiedler and Jain, 1992), which is only one component of biodiversity. Biodiversity is also more than species diversity (simply called diversity by some authors), which has been defined as the number of species in an area and their relative abundance (Pielou, 1977).

DeLong (1996) offered a more comprehensive definition:

> Biodiversity is an attribute of an area and specifically refers to the variety within and among living organisms, assemblages of living organisms, biotic communities, and biotic processes, whether naturally occurring or modified by humans. Biodiversity can be measured in terms of genetic diversity and the identity and number of different types of species, assemblages of species, biotic communities, and biotic processes, and the amount (e.g., abundance, biomass, cover, rate) and structure of each. It can be observed and measured at any spatial scale ranging from microsites and habitat patches to the entire biosphere.

This definition allows for modification according to the context in which it is used.

Various authors have proposed specific and detailed elaborations of this definition. Gaston and Spicer (1998) proposed a three-fold definition of “biodiversity”—ecological diversity, genetic diversity, and organismal diversity—while others conjoined the genetic and organismal components, leaving genetic diversity and ecological diversity as the principal components. These latter two elements can be linked to the two major “practical” value systems of direct use/genetics and indirect use/ecological described by Gaston and Spicer (1998). Other workers have emphasized a hierarchical approach or hierarchies of life systems.

In contrast, some argue that biodiversity, according to the definition of biological, does not include the diversity of abiotic components and processes, and that it is inaccurate to identify ecological processes, ecosystems, ecological complexes, and landscapes as components of biodiversity. The term ecological, as used in the sense of ecological system (ecosystem), encompasses both biotic and abiotic components and processes. Therefore, ecological diversity is a more appropriate term for definitions that include the diversity of ecological processes and ecosystems. However, ecological processes, it has been argued, should be included in the definition of biodiversity, the reasoning being that “although ecological processes are as much abiotic as biotic, they are crucial to maintaining biodiversity.” Similarly, a U.S. Bureau of Land Management advisory group included ecological processes in their definition of biodiversity in response to criticism that the Office of Technology Assessment's (1987) definition did not consider ecosystem form and function. Other writers point out that even though ecological processes are often cited as being crucial to maintaining biodiversity (Reid and Miller, 1989; Noss and Cooperrider, 1994; Samson and Knopf, 1994), this does not warrant the inclusion of ecological processes into the meaning of biodiversity. For example, Reid and Miller (1989) and Agarwal (1992) distinguished between biodiversity and the processes and ecological diversity that maintain it.

Nevertheless, the jargon word "biodiversity" is, by its very origin, fundamentally indefinable, being a populist word invented for convenience. Its invention has had beneficial effects by fuelling research projects, mainly in ecology and systematics, and scientists have been drawn into contributing to the debate by the need to show that biodiversity is useful to humans and necessary for the proper functioning of ecosystems. Conservation (i.e., management) of biodiversity is axiomatic to these two concerns and lies behind the scientific need to define the term within whatever context is appropriate, since no general definition will be suitable when applied across a range of situations.

Biodiversity conservation requires the management of natural resources, and this in turn requires the measurement of these resources. Biodiversity measurement implies the need for some quantitative value that can be ascribed to the various measurements so these values can be compared. Among the first scientists to measure diversity were Fisher, Corbet, and Williams (1943), who approximated the frequency distribution of the species represented by 1,2,3,4. . . (and so on) individuals by the logarithmic series αx, $\alpha x^2/2$, $\alpha x^3/3$, $\alpha x^4/4$. . . , where the constant α has been found to be a measure of species diversity. Species diversity is low when the number of species is growing slowly with respect to the increase in number of individuals, and it is high when the number of species is growing quickly.

If the need to quantify biodiversity drives the fundamental meaning of biodiversity, the definition may be limited to that which can be readily measured given current understanding and technologies. Such a definition of biodiversity could change over time as ideas, technology, and resources for measuring diversity change. DeLong (1996) suggested that an operational "clause" should be added to the definition of biodiversity, namely, that "biodiversity is. . .as measured in terms of. . . ." This approach provides a link to management while distinguishing between what biodiversity is (a state or attribute) and how it is measured. It also allows the operational clause to be adjusted over time without changing the fundamental meaning of the term. A definition of biodiversity should portray the full scope of what the term means, not just what can be measured and managed. In contrast, monitoring or management objectives must be attainable to be effective. Recognizing the distinction between a definition and management objectives should reduce the confusion between the meaning of biodiversity and the objectives for achieving biodiversity goals.

Biodiversity is a broad totality and often embraces elements beyond species diversity or numbers. For example, a major debate in biological sciences over many decades has been that of pattern versus process, especially in systematics and evolutionary studies. Molecular biology and systematics have enabled ecologists to see that inferred history is important in framing appropriate questions, and this understanding has precipitated a real integration of these twin hierarchies—pattern (e.g., diversity) and process (e.g., evolution). Fundamental divisions remain, such as "straight" parsimony (i.e., pattern) versus maximum likelihood (i.e., process) in the phylogenetic interpretation of sequence data.

It is apparent that the term biodiversity still lacks consistent meaning within the field of natural resource management. Michael Soulé found it shocking that "we are still trying to define biological diversity after all of the efforts of the Office of Technology Assessment and E. O. Wilson's book, *Biodiversity*" (Hudson, 1991). It is still defined in different ways by different people; some characterize biodiversity as being a widely used term "having no unified definition" and others emphasize or limit the meaning of biodiversity to that of native biodiversity. Some writers have included human alterations of biological communities in the scope of biodiversity (Bryant and Barber, 1994). Angermeier (1994) argued that "the absence of a 'native' criterion within the definition [of biodiversity] severely compromises biodiversity's utility as a meaningful biological concept," reasoning that native biodiversity is more valuable than artificial diversity and should be the primary focus of conservation efforts. The conservation of native biodiversity appears to be the theme of biodiversity conservation texts (Wilson and Peter, 1988; Hunter, 1996). Conversely, others argued that an important component of biodiversity is maintained by traditional farming techniques. In the context of conserving biodiversity, Reid and Miller (1989) and Bryant and Barber (1994) discussed the importance of genetic diversity within species of cultivated plants. Biodiversity within agricultural plants is important for pest management in agroecosystems and sustainable agriculture.

An accepted fundamental definition of biodiversity is needed for conservation planning, as are effective communication and co-operation within and among different countries, governments, agencies, disciplines, organizations, and private landowners. Co-operation among these entities has been identified as being necessary for the conservation of biodiversity (Babbitt, 1994). Knopf (1992) asserted that the definitions of biodiversity are "as diverse as the biological resource." Definitions of biodiversity range in scope from "the number of different species occurring in some location" to "all of the diversity and variability in nature" and "the variety of life and its processes." A more comprehensive definition is "the variety of living organisms, the genetic differences among them, the communities and ecosystems in which they occur, and the ecological and evolutionary processes that keep them functioning, yet ever changing and adapting" (Noss and Cooperrider, 1994).

This plethora of terms and definitions is one of the major stumbling blocks to reaching agreement in problem solving and decision making. If entities in a planning process view biodiversity in fundamentally different ways, agreement on management objectives and strategies for biodiversity conservation will be impaired. (Swingland, 1999).

The differences between these conceptual perspectives on the meaning of biodiversity, and the associated semantic problems, are not trivial. Management intended to maintain one facet of biodiversity will not necessarily maintain another. For example, a timber extraction program that is designed to conserve biodiversity in the sense of site species richness may well reduce biodiversity measured as genetic variation within the tree species harvested. Clearly, the maintenance of different facets of biodiversity will require different management strategies and resources, and will meet different human needs.

Even if complete knowledge of particular areas could be assumed, and standard definitions of diversity are derived, the ranking of such areas in terms of their importance with respect to biological diversity remains problematic. Much depends on the scale that is being used. Thus, the question of what contribution a given area makes to *global* biological diversity is very different from the question of what contribution it makes to local, national, or regional biological diversity. This is because, even using a relatively simplified measure, any given area contributes to biological diversity in at least three different ways—through its richness in numbers of species, through the endemism (or geographical uniqueness) of these species (e.g., Mittermeier *et al.*, 1992), and on the basis of degree of threat. The relative importance of these three factors will inevitably change at different geographical scales, and sites of high regional importance may have little significance at a global level. None of these factors includes any explicit assessment of genetic diversity.

Although the word biodiversity has already gained wide currency in the absence of a clear and unique meaning, greater precision will be required of its users if policy and programs are to be more effectively defined in the future.

III. GENETIC DIVERSITY

Genetic diversity is reliant on the heritable variation within and between populations of organisms. New genetic variation arises in individuals by gene and chromosome mutations, and in organisms with sexual reproduction it can be spread through the population by recombination. It has been estimated that in humans and fruit flies alike, the number of possible combinations of different forms of each gene sequence exceeds the number of atoms in the universe. Other kinds of genetic diversity can be identified at all levels of organization, including the amount of DNA per cell and chromosome structure and number. Selection acts on this pool of genetic variation present within an interbreeding population. Differential survival results in changes of the frequency of genes within this pool, and this is equivalent to population evolution. Genetic variation enables both natural evolutionary change and artificial selective breeding to occur (Thomas, 1992).

Only a small fraction ($<1\%$) of the genetic material of higher organisms is outwardly expressed in the form and function of the organism; the purpose of the remaining DNA and the significance of any variation within it are unclear (Thomas, 1992). Each of the estimated 10^9 different genes distributed across the world's biota does not make an identical contribution to overall genetic diversity. In particular, those genes that control fundamental biochemical processes are strongly conserved across different taxa and generally show little variation, although such variation that does exist may exert a strong effect on the viability of the organism; the converse is true of other genes. A large amount of molecular variation in the mammalian immune system, for example, is possible on the basis of a small number of inherited genes (Thomas, 1992).

IV. SPECIES DIVERSITY

Historically, species are the fundamental descriptive units of the living world and this is why biodiversity is very commonly, and incorrectly, used as a synonym of *species diversity,* in particular of "species richness," which is the number of species in a site or habitat. Discussion of global biodiversity is typically presented in terms of global numbers of species in different taxonomic groups. An estimated 1.7 million species have been described to date; estimates for the total number of species existing on earth at present vary from 5 million to nearly 100 million. A conservative working estimate suggests there might be around 12.5 million.

When considering species numbers alone, life on earth appears to consist mostly of insects and microorganisms. The species level is generally regarded as the most natural one at which to consider whole-organism diversity. While species are also the primary focus of evolutionary mechanisms, and the origination and extinction of species are the principal agents in governing biological diversity, species cannot be recognized and enumerated by systematists with total precision. The concept of what a species is differs considerably among groups of organisms. It is for this reason, among others, that species diversity alone is not a satisfactory basis on which to define biodiversity.

Another reason why a straightforward count of the number of species provides only a partial indication of biological diversity concerns the concept of degree or extent of variation that is implicit within the term biodiversity. By definition, organisms that differ widely from each other in some respect contribute more to overall diversity than those that are very similar. The greater the interspecific differences (e.g., by an isolated position within the taxonomic hierarchy), then the greater contribution to any overall measure of global biological diversity. Thus, the two species of Tuatara (genus *Sphenodon*) in New Zealand, which are the only extant members of the reptile order Rhynchocephalia, are more important in this sense than members of some highly species-rich family of lizards. A site with many different higher taxa present can be said to possess more *taxonomic diversity* than another site with fewer higher taxa but many more species. Marine habitats frequently have more different phyla but fewer species than terrestrial habitats, that is, higher taxonomic diversity but lower species diversity. By this measure, the Bunaken reef off the north coast of Sulawesi has the highest biodiversity on earth. Current work is attempting to incorporate quantification of the evolutionary uniqueness of species into species-based measures of biodiversity.

The ecological importance of a species can have a direct effect on community structure, and thus on overall biological diversity. For example, a species of tropical rain forest tree that supports an endemic invertebrate fauna of a hundred species makes a greater contribution to the maintenance of global biological diversity than does a European alpine plant that may have no other species wholly dependent on it.

V. ECOSYSTEM DIVERSITY

While it is possible to define what is in principle meant by genetic and species diversity, it is difficult to make a quantitative assessment of diversity at the ecosystem, habitat, or community level. There is no unique definition or classification of ecosystems at the global level, and it is difficult in practice to assess ecosystem diversity other than on a local or regional basis, and then only largely in terms of vegetation. Ecosystems are further divorced from genes and species in that they explicitly include abiotic components, being partly determined by soil/parent material and climate.

To get around this difficulty, ecosystem diversity is often evaluated through measures of the diversity of the component species. This may involve assessment of the relative abundance of different species as well as consideration of the types of species. The more that species are equally abundant, then the more diverse that area or habitat. Weight is given to the numbers of species in different size classes, at different trophic levels, or in different taxonomic groups. Thus a hypothetical ecosystem consisting only of several plant species would be less diverse than one with the same number of species but that included animal herbivores and predators. Because different weightings can be given to these different factors when estimating the diversity of particular areas, there is no one authoritative index for measuring ecosystem diversity. This obviously has important implications for the conservation ranking of different areas. In examining beta diversity (i.e., the change in species composition between areas), the only reliable predictor of community similarity is to compare the species composition of the site immediately adjacent.

VI. BIODIVERSITY: MEANING AND MEASUREMENT

A. Species Diversity

A. S. Corbet, upon analyzing a large collection of butterflies from Malaya, remarked on the decrease in number of new species with an increasing number of individuals. He thought that the resulting distribution could be described by a hyperbola, but R. A. Fisher, to whom Corbet sent his results, suggested that a negative binomial distribution would be much more appropriate

(Williams, 1964). As mentioned earlier, Fisher, Corbet, and Williams (1943) approximated the frequency distribution of the species represented by 1,2,3,4. . . (and so on) individuals by the logarithmic series αx, $\alpha x^2/2$, $\alpha x^3/3$, $\alpha x^4/4$. . . , where the constant α is a measure of species diversity. Species diversity is low when the number of species rises slowly with an increase in the number of individuals, and diversity is high when the number of species rises quickly.

Species diversity measurement was thus clearly formulated more than 50 years ago and a particular index was proposed. Fisher *et al.* attempted to find some general "rule" or "law" according to which the numerical abundances of different species were related to each other. In many communities, the number of species with given abundance could be approximated by the log-normal distribution. If species are classified in accordance with their abundance in logarithmically increasing classes—so-called "octaves" (i.e., the first octave contains 1–2 individuals, the second contains 2–4 individuals, the third has 4–8, the fourth has 8–16, and so on)—then the number of species per "octave" shows a truncated normal distribution. If a sample contains a high number of species and individuals, we can usually obtain a log-normal distribution, and it is obviously more tractable than the logarithmic series.

MacArthur (1957) went further by proposing an interesting model that assumed that boundaries between niches in resource–niche hypervolume are set at random, whereas the relative abundances of species are proportional to these sections of hypervolume. This model became widely known as the "broken-stick" or MacArthur's model. The distribution of abundance prescribed by MacArthur's model is much "flatter" (i.e., the contrast between given species and the next in the sequence is less) than in the case of a logarithmic series (Ghilarov, 1996).

It has become clear that there is no universal type of distribution of relative abundance that corresponds to all real communities, though such distributions change in the course of succession according to a particular pattern. The dominance of a few of the most abundant species is more pronounced at the early stages of succession, while later the species of intermediate abundance become more significant (Whittaker, 1972). A comprehensive understanding of the underlying mechanisms that result in a given pattern of species abundance still eludes scientists.

Another line of species diversity studies was connected with the use of special indices proposed to measure diversity without reference to some hypothetical distribution of relative abundance. A great variety of indices were proposed that assess the number of species and the proportions in abundance of different species. Among others, there was the very popular index that is based on Shannon's formula derived from information theory:

$$H = \Sigma\, p_i \log p_i$$

where p_i is the proportion of the total number of individuals that belong to the *i*th species.

In a seminal work on the measurement of diversity, Whittaker (1972) introduced the concepts of alpha, beta, and gamma diversity. The measurements just described, giving diversity values for single sites, are examples of alpha diversity. The beta and gamma diversity concepts relate to changes in diversity between sites at local (beta) and geographical (gamma) scales. An essential part of these relational concepts is the idea of species turnover—the degree to which species replace other species at different sites. For use in assessing the relative value of multiple sites for the conservation of biodiversity, the idea of species turnover is translated into the principle of complementarity (see Section VIII,A), which can be implemented in combination with a taxonomic diversity index.

B. Taxonomic Diversity

Biodiversity measurements that measure genetic difference directly, or indirectly through use of the taxonomic (cladistic) hierarchy (Williams *et al.,* 1991), are currently being used. The indirect taxonomic approach is more practical because we already have a "rule of thumb" taxonomic hierarchy (which is being steadily improved through the application of cladistic analysis, notably to molecular data), whereas reliable estimates of overall genetic differences between taxa are virtually non-existent (abridged from Vane-Wright, 1992).

Based on the shared and unshared nodes between taxa (equivalent to position in the taxonomic hierarchy), a number of taxonomic diversity indices have now been developed. Of these, the most distinct are root weight, higher taxon richness, and taxonomic dispersion. The first places highest individual value on taxa that separate closest to the root of the cladogram and comprise only one or relatively few species; in effect this gives high weighting to relict groups (Vane-Wright, 1996). Higher taxon richness favors taxa according to their rank and number of included species. Dispersion, the most complex of the measures proposed so far (Williams *et al.,* 1991), endeavors to select an even spread of taxa across the hierarchy, sampling a mixture of high, low, and intermediate ranking groups.

For a given group these measures, together with simple species richness if desired, can be used to compare the biotic diversity of any number of sites. The measures can also be expressed as percentages. Thus a site with viable populations of all species in a group would have a diversity score of 100%, whereas a site without any species of the group in question would score zero. In reality, of course, most sites have only a selection of species, and so receive various intermediate scores. Such assessments allow us to compare all sites with each other, and rank them individually from highest to lowest diversity (Vane-Wright, 1996). However, if we then take some conservation action (such as conserving a particular site), the same measures are unlikely to be directly comparable for making a second decision (such as choosing a second conservation site). This is because, in most real situations at least, there will be considerable overlap in the presence of species at particular sites.

C. Community Diversity

Early ecologists did not confine themselves to measuring species diversity. They also tried to understand the relationship of diversity with other features of the community (e.g., Williams, 1964; Whittaker, 1972). The dependence of species diversity on the structural complexity of the environment was demonstrated (MacArthur and MacArthur, 1961), as was the role of predation (Addicott, 1974) and periodical disturbance (Sousa, 1979) in determining a given level of diversity. The relationship between the species diversity and standing crop of a community was also shown (Ghilarov and Timonin, 1972).

Margalef (1957) was the first to use the Shannon index (though expressed in a different form). He proposed to evaluate the level of community organization in terms of information theory. Margalef stimulated many ecologists to quantitatively measure the species diversity of different communities and/or of the same community in different stages of its development. At that time, there was a widespread belief that with a single numerical value, an assessment could be made of some very significant feature of community structure. Many ecologists believed that in measuring species diversity at the community level they were using an approach that was fundamental to an understanding of diversity (Ghilarov, 1996).

Ecologists have measured diversity either by estimating species richness (number of species) in an area, or by using one or more indices combining species richness and relative abundance within an area. Some attempts have also been made to measure change in species richness (species turnover) between areas. These solutions to the problem of measuring biodiversity are limited because species richness takes no account of the differences between species in relation to their place in the natural hierarchy. Moreover, relative abundance is not a fixed property of a species, for it varies widely from time to time and place to place. In many environments most taxa are virtually or even completely unknown.

Conservation biologists, or applied ecologists, have called for a measurement of diversity that is more clearly related to overall genetic difference. An example concerns the problem of differential extinction. In *World Conservation Strategy* (IUCN/UNEP/WWF, 1980), it is noted that "the size of the potential genetic loss is related to the taxonomic hierarchy because. . .different positions in this hierarchy reflect greater or lesser degrees of genetic difference. . . . The current taxonomic hierarchy provides the only convenient rule of thumb for determining the relative size of a potential loss of genetic material."

D. Synthesis

A model incorporating island biogeographic theory, species abundance, and speciation, and that produces a fundamental biodiversity number (θ) that is closely associated with species richness and abundance in an equilibrium meta-population, has been proposed in Hubbell's unified theory (1997). This model assumes zerosum community dynamics or a saturated, totally stochastic local community, which limits its application, but it advances the study of species richness and relative abundance if others can extend its usefulness to the nonequilibrium systems that characterize the real world.

VII. BIODIVERSITY: CHANGES IN TIME AND SPACE

A. Changes Over Time

The fossil record is very incomplete, which emphasizes the marked variation between higher taxa and between species in different ecosystems in the extent to which individuals are susceptible to preservation and subsequent discovery. Chance discovery has played a large part in compiling the known fossil record, and interpretation by paleontologists of the available material is beset by differences of opinion. Thus, the record is relatively good for shallow-water, hard-bodied marine invertebrates, but poor for most other groups, such as plants in moist tropical uplands.

Two relevant points appear to be well substantiated. First, *taxonomic diversity,* as measured by the number of recognized phyla of organisms, was greater in Cambrian times than in any later period. Second, it appears that *species diversity* and the number of families have undergone a net increase between the Cambrian and Pleistocene epochs, although interrupted by isolated phases of mass extinction (few of which are reflected in the fossil record of plants).

B. Changes in Space

Species diversity in natural habitats is high in warm areas and decreases with increasing latitude and altitude; additionally, terrestrial diversity is usually higher in areas of high rainfall and lower in drier areas. The richest areas are tropical moist forest and, if current estimates of the number of microfaunal species (mainly insects) of tropical moist forests are credible, then these areas, which cover perhaps 7% of the world's surface area, may well contain over 90% of all species. If the diversity of larger organisms only is considered, then coral reefs such as Bunaken (see earlier) and, for plants at least, areas with a Mediterranean climate in South Africa and Western Australia may be as diverse. Gross genetic diversity and ecosystem diversity will tend to be positively correlated with species diversity.

What are not fully understood are the reasons for the large-scale geographic variation in species diversity, and in particular for the very high species diversity of tropical moist forests. The origin of diversity through the evolution of species and the maintenance of this diversity both need more study before they are better understood. This will require consideration of the present and historic (in a geological or evolutionary sense) conditions prevailing in particular areas, principally climatic but also edaphic and topographic. Climatically benign conditions (warmth, moisture, and relative aseasonality) over long periods of time appear to be particularly important.

Climax ecosystems will be more diverse than areas at earlier successional stages, but an area with a mosaic of systems at different successional stages will probably be more diverse than the same area at climax provided that each system occupies a sufficiently large area of its own. In many instances, human activities artificially maintain ecosystems at lower successional stages. In areas that have been under human influence for extended periods, notably in temperate regions, maintenance of existing levels of diversity may involve the maintenance of at least partially man-made landscapes and ecosystems, mixed with adequately sized areas of natural climax ecosystems.

VIII. LOSS OF BIODIVERSITY AND CAUSES

Species extinction is a natural process that occurs without the intervention of humans since, over geological time, all species have a finite span of existence. Extinctions caused directly or indirectly by humans are occurring at a rate that far exceeds any reasonable estimates of background extinction rates, and to the extent that these extinctions are correlated with habitat perturbation, they must be increasing.

Quantifying rates of species extinction is difficult and predicting future rates with precision is impossible. The documentation of definite species extinctions is only realistic under a relatively limited set of circumstances, for example, where a described species is readily visible and has a well-defined range that can be surveyed repeatedly. Unsurprisingly, most documented extinctions are of species that are easy to record and that inhabit sites that can be relatively easily inventoried. The large number of extinct species on oceanic islands is not solely an artifact of recording, because island species are generally more prone to extinction as a result of human actions.

Most global extinction rates are derived from extrapolations of measured and predicted rates of habitat loss, and estimates of species richness in different habitats. These two estimates are interpreted in the light of a principle derived from island biogeography, which states that the size of an area and of its species complement tend to have a predictable relationship. Fewer species are able to persist in a number of small habitat fragments than in the original unfragmented habitat, and this can result in the extinction of species (MacArthur and Wilson, 1967). These estimates involve large degrees of uncertainty, and predictions of current and future extinction rates should be interpreted with considerable caution. The pursuit of increased accuracy in the estimation of global extinction rates is not crucial. It is more important to recognize in general terms the extent to which populations and species that are not monitored are likely to be subject to fragmentation and extinction (Temple, 1986).

Loss of biodiversity in the form of domesticated animal breeds and plant varieties is of little significance in terms of overall global diversity, but genetic erosion in these populations is of particular human concern in so far as it has implications for food supply and the sustainability of locally adapted agricultural practices. For domesticated populations, the loss of wild relatives of crop or timber plants is of special concern for the same reason. These genetic resources may not only underlie the productivity of local agricultural systems but may also, when incorporated into breeding programs, provide the foundation of traits (disease resistance, nutritional value, hardiness, etc.) that are of global importance in intensive systems and that will assume even greater importance in the context of future climate change. Erosion of diversity in crop gene pools is difficult to demonstrate quantitatively, but can be indirectly assessed in terms of the increasing proportion of world cropland planted to high-yielding, but genetically uniform, varieties. Genetic modification of organisms, varieties, or cultivars for food production, pharmaceuticals, and other products, which has caused concern in some countries but not others, may also contribute to the loss of biodiversity.

Humans exterminate species either directly by hunting, collection, and persecution or indirectly through habitat destruction and modification. Overhunting is perhaps the most obvious direct cause of extinction in animals, but it is undoubtedly far less important than the indirect causes of habitat modification in terms of overall loss of biodiversity. Hunting selectively affects the targeted species, as well as plant and animal species whose populations are subsequently affected either negatively or positively, and so it has important implications for the management of natural resources. Genetic diversity in a hunted population is liable to decrease as a result of the same factors. The genetic diversity represented by populations of crop plants or livestock is also likely to decline as a result of mass production, for the desired economics of scale demand high levels of uniformity.

Sustained human activity will affect the relative abundance of species and in extreme cases may lead to extinction. This may result from the habitat being made unsuitable for the species (e.g., clear-felling of forests or severe pollution of rivers) or through the habitat becoming fragmented (discussed earlier). Fragmentation divides previously contiguous populations of species into small sub-populations. If these are sufficiently small, then chance processes lead to higher probabilities of extinction within a relatively short time. Major changes in natural environments are likely to occur within the next century as a result of changes in global climate and weather patterns. These will cause greatly elevated extinction rates.

IX. MAINTAINING BIODIVERSITY

A. *In Situ* Conservation

The maintenance of biological diversity is the sustainable management of viable populations of species or populations *in situ* or *ex situ*. The maintenance of a significant proportion of the world's biological diversity only appears feasible by maintaining organisms in their wild state and within their existing range. This allows for continuing adaptation of wild populations by natural evolutionary processes and, in principle, for current utilization practices to continue. For such maintenance to succeed, it almost invariably requires enhanced management through the integrated, community-based conservation of protected areas.

Over the last thirty years, conservation biologists have struggled with the concept of the maintenance of biodiversity in highly diverse environments like rain forests. Analytical techniques (neural-net models) that allow us to reconstruct past distributions of forest types present an opportunity to predict past contractions and expansions of forest forms, and the likelihood of refugia surviving climate change. Such extrapolations must be treated with caution, as pollen samples from Brazil (for example) disproved modeling predictions that savanna grasslands should have been extant, when in fact tropical and temperate forests were present. Various authors also opposed the Pleistocene refugia hypothesis (Haffer, 1969) for the Amazon region because some evidence demonstrated the lack of rain forest fragmentation during that era. In the biogeographical zones of the Australian wet tropics, there is a strong correlation between diversity patterns and reputed rain forest refugia in both species and genetic diversity. However, this appears to have been caused by differential extinction rates in differently sized refugia rather than by allopatric speciation in the Pleistocene. Others have emphasized that a greater concentration on the Pliocene or before would be useful, since most tropical species radiations occurred before the Pleistocene.

The local-determination hypothesis of species diversity (Rosenzweig, 1995), which predicts similar species diversity in similar habitats, has also been challenged. In sister taxa of plants, the net diversification was significantly higher in Asia than in North America for genera shared between the two continents. Greater insights into the effects of current ecology on the local diversity of an area may be assisted by considering the relative ages of clades, which could establish species proliferation rates between regions, thus advancing the local versus regional diversity debate (Ricklefs and Schluter, 1993). They also tested the taxon cycle theory (Wilson, 1961) using phylogenies of bird species and showed that older species' lineages had more restricted ranges, smaller habitat breadth, and more fragmented distributions, and were closer to extinction than younger species.

In efforts to conserve biodiversity, preserving genetic dissimilarity is often a higher priority than maintaining genes of considerable similarity. Recent work shows that genetic divergence in mammals increases from the headwaters to the mouth as a river gets broader and thus becomes a greater barrier to populations on opposite banks; this effect promotes species diversity through allopatric speciation. Headwater species are basal in the phylogeny, and shared haplotypes occur only at the headwaters; this research is a contribution to Wallace's riverine diversification hypothesis in the Amazon basin.

A central question in the design of effective conservation programs is what geographical regions to protect in order to maintain the most biological diversity. The term biodiversity hotspot was coined by Myers (Myers, 1990) and most commonly refers to regions of high species richness. GAP analysis is used to identify gaps in existing protected area networks (Scott *et al.*, 1993); it uses algorithms to select the minimum set of grid cells that encompass the unprotected species. Rarity and endemicity have been used to define hotspots in bird conservation (Balmford and Long, 1994), and species richness and endemism have been used to rank countries (McNeely *et al.*, 1990). Hotspots are also defined as those areas with the greatest number of threatened species.

In setting conservation priorities, assumptions are made that indicator groups (e.g., macro-organisms such as birds, mammals, and plants) are good predictors of biological diversity in general. Another question that arises is how best to analyze biodiversity information to generate accurate and useful analyses that will inform conservation decisions. On a large scale, some concordance is found between bird diversity across continents with insect diversity (Pearson and Cassola, 1992), and in endemism patterns across taxa (Lawton, 1994); but at a finer spatial scale this correlation begins to break down. Richness in genera and families are good predictors of species richness at a finer level (Balmford *et al.*, 1996a, 1996b). However, species richness is not a good measure with which to identify hotspots for conservation because it overlooks rare species, although as the sample area for hotspots is increased, more rare species are included as a simple function of arithmetic progression. Rarity and endemicity are efficient indices for selecting the most parsimonious number of sites, but compared to complementarity measures they are less useful in defining conservation priorities.

A good conservation measure is complementarity, where the species complement of a reserve or area is identified and then further sites are found that add the greatest number of new species; this is akin to the portfolio approach (Swingland, 1997). Another method using integer linear programming to choose the optimal set of sites (maximal-covering-location; Church *et al.*, 1996) is limited to small datasets and does not achieve the greatest conservation gain for the fewest additional sites. Clearly, combining an ecosystem portfolio approach with a richness or endemism assessment would be effective, but differing approaches are needed according to the conservation goal and data availability.

B. *Ex Situ* Conservation

Viable populations of many organisms can be maintained in cultivation or in captivity. Plants may also be maintained in seed banks and germplasm collections; similar techniques are under development for animals (storage of embryos, eggs, and sperm, i.e., "frozen zoos") but are more problematic. *Ex situ* conservation is extremely costly in the case of most animals, and while it would in principle be possible to conserve a very large proportion of higher plants *ex situ,* this would be feasible for only a small percentage of the world's organisms. Furthermore, it often involves a loss of genetic diversity through founder effects and the high probability of inbreeding (Milner-Gulland and Mace, 1998).

X. CONTEXTUAL VARIATIONS OF THE DEFINITION

A. Derivation of "Biodiversity"

The definition of biodiversity put forth by the Office of Technology Assessment (1987) appears to be the most widely cited basis for other published definitions

(Scott *et al.,* 1995). However, the OTA did not explain why they defined the term as they did, nor did they cite any supportive documentation. One problem with relying solely on authoritative sources for definitions of biodiversity is that different authorities have defined the term in fundamentally different ways.

"Bio" is derived from the Greek word *bios,* meaning life. Biological and biotic are terms that refer to life, living organisms, assemblages of living organisms, and the activities and interactions of living organisms. The scope of the term biological can be further understood in the context of components and processes that are considered biological. Defining biodiversity (i.e., diversity) is more difficult because it continues to be defined in several fundamentally different ways. In definitions of biodiversity, diversity has been characterized as (1) the number of different types of items, (2) the number of different types of items and their relative abundance, and (3) variety. Characterization of diversity in discussions of biodiversity has also included the structural complexity of landscapes (Huston, 1994).

B. Classifying Biodiversity

The classification of biodiversity can be divided into those authors who consider biodiversity to be a state and those who believe that it is a measure of the state.

Most authors have defined biodiversity as a state or attribute, for example, "biodiversity is the variety of. . ." or "variety and variability of. . ." (Noss and Cooperrider, 1994). Standard dictionaries have classified diversity as a state, condition, or quality (Soukhanov *et al.,* 1988).

Other definitions of biodiversity limited the scope of the attribute to explicit, quantifiable dimensions or measures, for example, "biodiversity is the number of. . ." or "the number and relative abundance of. . ." (Office of Technology Assessment, 1987). This emphasis on quantitative, operational definitions of biodiversity and criticisms of non-quantitative definitions (Angermeier, 1994; Hunter, 1996) may signal a potential shift in the classification of the term from an attribute to a measure of an attribute. In the ecological and natural resource management literature, Pielou (1977) and others have treated diversity as a one- or two-dimensional attribute of a community (e.g., diversity is "the number of" or "the number and relative abundance of"). More recently, it has been defined as a measure or index of those attributes; for example, diversity is a "measure of. . ." (Noss and Cooperrider, 1994). Operational definitions of biodiversity (Angermeier, 1994) provide impetus to define biodiversity in quantitative terms as Hunter (1996) recommended.

C. Attributes of Biodiversity

Another way of delineating the meaning of a term is to list its characteristics, properties, qualities, and parts. Noss (1990) recognized three main attributes of biodiversity: composition, structure, and function.

Composition addresses the identity and richness of biotic components, and the relative amount (e.g., abundance, cover, biomass) of each (Noss, 1990). Biotic components of ecosystems include genes, organisms, family units, populations, age classes, species and other taxonomic categories, trophic levels of animals (e.g., herbivores, predators), animal guilds and assemblages, plant communities, and interacting assemblages of plants, animals, and microorganisms (i.e., biotic communities).

Structural attributes of biodiversity refer to the various vertical and horizontal components of a community or landscape (Noss, 1990) and the organizational levels of plant and animal populations and assemblages (Gaston and Spicer, 1998; Hunter, 1996). Considering only biotic, vegetative components of a landscape, horizontal structure consists of the size, shape, and spatial arrangement and juxtaposition of different plant communities; vertical structure consists of the foliage density and height of different vegetation layers (Noss, 1990). Structure can also refer to population, age and trophic structure, and other levels of community organization (Hunter, 1996).

The inclusion of structure in the meaning of biodiversity provides linkages with other concepts, such as habitat diversity and the plant community concept, for both of which vegetation structure is an important differentiating attribute. Structure may have been left out of most definitions of biodiversity because the concept of biodiversity evolved from the concept of ecological diversity, which primarily focused on species diversity (Fisher *et al.,* 1943). Interestingly, 20 years ago it was asserted that measurements of diversity should not preclude structural diversity even though the term is most often used in reference to species diversity. Diversity can also be used in reference to niche width and the structural complexity of habitats.

Biotic functions represent the third component of biodiversity, and these include processes such as herbivory, predation, parasitism, mortality, production, vegetative succession, nutrient cycling and energy flow through biotic communities, colonization and extinction, genetic drift, and mutation (Noss, 1990). Biotic processes can be addressed in terms of the identity and number of different types of processes, as well as the rate (e.g., predation rate) at which each process operates.

Diversity of biotic components and processes can be observed at many biogeographic scales, from microsites and larger-scale landscape elements (e.g., vegetation types, habitat types, range sites) to regional landscapes, biomes, continents, hemispheres, and the entire biosphere (Noss, 1990; Huston, 1994; Hunter, 1996). Although these are scales at which biodiversity can be observed, they are not necessarily scales of biodiversity because most include abiotic (e.g., geological) features. Biodiversity can also be observed at several organism-based scales, including individual organisms, populations, species, and assemblages (e.g., guilds and plant communities), which themselves can be observed at various biogeographical scales.

D. Biological Resource Asset and Management Objectives

The contextual variations in the definition of biodiversity depend on what use is being made of the biological resource asset (or bioasset), and thus the asset management objective. Biological resource values consist of direct use, indirect use, and option and non-use values. For the purposes of assessing potential use, they can be further classified as follows:

- Direct use values of *major extractive products.* Principally, this would include forestry for timber and commercial fisheries in the case of terrestrial and marine systems. Extraction of these products often involves substantial investment in capital equipment by large non-local firms, and the products are transported and sold in well-developed markets far from their original source.
- Direct use values of *"minor" extractive products.* These are naturally or semi-naturally occurring products that require labor-intensive gathering or harvesting activities, often carried out by local people. Examples include rattan, fuelwood, seaweed, wild foods, artisanal fisheries, aquarium fish, and medicinal herbs. These may be collected for sale, barter, or home consumption.
- Direct use values that require the extraction of only a small amount of biological material for *ex situ research* or storage. This includes extraction of material for biological inventories, germplasm banks, and industrial research. Extraction is often accomplished during short or long expeditions that traverse large areas to collect representative samples of biological material.
- Direct use values that are *non-extractive,* but often require considerable on-site interaction of the user with the resource. This includes ecotourism, recreation, on-site research, and other major "non-consumptive" activities occurring principally in protected areas. These activities are characterized by the need to provide food, lodging, and transport to the participants.
- Indirect use values that accrue *on site.* The primary feature of these values is that they support or protect the basic functioning of the protected area. Examples include nutrient cycling, stabilization of soils in erosion-prone areas, coastal zone stabilization, and biological support to local ecosystems. As a result of their nature, the value of these on-site functions is likely to be a component of all the other direct and non-use values generated by the area.
- Indirect use values that accrue *off site.* The value of these functions—such as watershed protection, natural ecosystems protected as national parks in generating income from wildlife tourism, protection of fisheries' nurseries and subsistence fisheries, and climate regulation—may be very large or very small depending on their relative importance to the support or protection of off-site economic activity.
- Option values. Because option values may be associated with each and every use value, they are considered only where they may be of potential significance in conjunction with the particular type of product or service.
- Nonuse values. By their very nature these values occur at a distance from the resource and require no extraction or physical interaction with the resource, for example, stewardship, ethics, cultural belief, and aesthetics.

These foregoing values are only indirectly related to biological diversity. That is, a certain level of species richness is required for these functions but there is not necessarily a direct correlation between the value of the ecosystem and its diversity. Thus, mangrove ecosystems are generally of far lower diversity than adjacent lowland terrestrial forests, but in resource terms they are likely to be of comparable value. The savannas of eastern and southern Africa, which are of great importance in generating revenues from tourism, are less diverse than the moist forests in these countries, which have far less potential for tourism.

E. *Cave Canem* or the Precautionary Principle

At present, humans actively exploit a relatively small proportion of the world's biological diversity. Many other potential, yet undiscovered, optional and non-use values of biodiversity exist. These factors support a precautionary approach to maintaining biological diversity. In this case, the precautionary principle argues that actions should be taken to prevent further loss of biodiversity and potentially irreversible consequences *before* all biological uncertainties are resolved. Yet in conserving biodiversity, there must come a point at which the projected costs required to protect and maintain it will outweigh any probable benefits.

If species are to be viewed as a resource, and their maintenance is to be cost-effective, conservation should concentrate on systems and areas rich in species, and on those species known to be useful, or regarded as having a high probability of being useful. Thus biodiversity and its conservation would be defined purely along operational or cost–benefit lines. This bioasset perspective on biodiversity would therefore rest upon economic arguments more than biological ones.

Biodiversity has been identified as important for ecosystem health, medicinal values, agricultural purposes, and aesthetic and recreational values (Noss and Cooperrider, 1994). Noss (1990) characterized an operational definition as one that is responsive to real-life management and regulatory questions, adding that such a definition is unlikely to be found for biodiversity. Angermeier (1994) referred to an operational definition in a similar way, and Hunter (1996) suggested that a quantitative definition is needed for monitoring biodiversity and developing management plans. On the other hand, some writers assert that the confounding of definition and application is partly to blame for the confusion over how biodiversity concepts can be practically implemented.

XI. IMPLICATIONS OF VARIATIONS IN THE DEFINITION

The need for an unequivocal and precise meaning of biodiversity that is scientifically sensible and universally applicable is imperative to help guide the design of policy and programs for the future, as well as to make critical decisions in the present. Currently, such a definition does not exist. As a concept, biodiversity is both ubiquitous and useful, particular and confusing; and for this reason it is constantly redefined on nearly every occasion.

One of the many reasons for this state of affairs is that the definition of biodiversity affects objectives in national and regional research and conservation management, and in international funding priorities. One could easily promote a timber extraction or non-timber forest product program that conserves species richness (i.e., numbers of species) at the expense of genetic diversity. Indeed, a current research program to stimulate or increase the range of tropical tree species not currently in trade, as a way to take the pressure off over-exploited species, may be misguided. It may lead to increased genetic as well as species impoverishment when foresters expand the number of species they take and select only the best and most mature specimens, thus removing the most productive and healthiest genetic stock.

Apart from the principal definitions of biodiversity discussed earlier, such as the highest number of species (i.e., species richness) and the highest level of species endemicity (Myers, 1990) or taxal endemicity (called critical faunas analysis), interpretations of pure or applied definitions are becoming more common within the vocabulary of conservation and biodiversity utilization when determining biodiversity management priorities. Some examples are national biodiversity programs that maintain "biodiversity portfolios"; biodiversity defined as flagship or keystone species diversity; viability modeling (population viability analysis) defining the species' populations to be prioritized; population analysis defining sustainability and thus defining a species' status; projects that focus on the feasibility of integrating the targeted species, assemblages, or ecosystems with the needs of local human populations and sustainable use; and (lastly) political exigency (Swingland, 1997). Although the conservation policy of a country may be driven by more pressing needs—family planning, education, politics, internal conflict, financial planning and investment, individual vested interests—current policy and decisions are also being made on the foregoing biodiversity bases rather than along strict academic lines.

Endemicity and species richness are useful starting points in defining priorities on the global level, but without information on the possibility of extinction using viability modeling or population analysis, the urgency of a given conservation action cannot be assessed. Moreover, with the increasing emphasis on the integration of local people into conservation programs to minimize long-term costs and to provide a more stable basis for the people and their natural environment, the potential for community-based conservation, coupled with sustainable use, cannot be ignored. Since national or external funding will generally provide essential support for most projects, the ecological importance of an area relative to others using an ecosystem diversity (or portfolio) approach will be a major selection criterion. The presence of a flagship or keystone species will also be significant in raising such funds. Clearly political exigencies or pure chance can enter the situation, and scientists have yet to articulate whether genetic diversity should be used as the key measure. In the absence of realistic methods of quantifying these biodiversity characteristics, they must remain imponderable objectives for the moment.

The differing approaches being advocated for biodiversity conservation are not just guided by the available methodologies but are also symptomatic of the underlying philosophies. The evolution-based approach is predominantly the preserve of biologists, and it is concerned with the maintenance of diversity as an unqualified objective unaffected by economics. The need for conservation and the uses of biodiversity—the resource-based argument—are what are used to "sell" the proposition to decision makers and policy-makers. Where these factors come together, the ideal of ecological sustainability and the conservation methods of achieving it will be possible. Because so much is now formally invested in using the word biodiversity, its definition will continue to play a crucial role in both conservation planning and public policy.

See Also the Following Articles

BIODIVERSITY, ORIGIN OF • ECOLOGY, CONCEPT AND THEORIES OF • ECOSYSTEM, CONCEPT OF • GENETIC DIVERSITY • HABITAT AND NICHE, CONCEPT OF • LOSS OF BIODIVERSITY, OVERVIEW • MEASUREMENT AND ANALYSIS OF BIODIVERSITY • TAXONOMY, METHODS OF

Bibliography

Addicott, J. F. (1974). Predation and prey community structure: An experimental study of mosquito larvae on the protozoan communities of pitcher plants. *Ecology* **55**, 475–492.

Agarwal, A. (1992). Sociological and political constraints to biodiversity conservation: A case study from India. In *Conservation of Biodiversity for Sustainable Development* (O. Sandlund, K. Hindar, and A. H. D. Brown, eds.), pp. 293–302. Scandinavian University Press, Oslo, Norway.

Angermeier, P. L. (1994). Does biodiversity include artificial diversity? *Conservation Biol.* **8**, 600–602.

Babbitt, B. (1994). Protecting biodiversity. *Nature Conservancy* **44**, 1621.

Balmford, A., and Long, A. (1994). Avian endemism and forest loss. *Nature* **372**, 623–624.

Balmford, A., Green, M. J. B., and Murray, M. G. (1996a). Using higher-taxon richness as a surrogate for species richness: I. Regional tests. *Proc. Roy. Soc. London, Ser. B* **263**, 1267–1274.

Balmford, A., Jayasuriya, A. H. M., and Green, M. J. B. (1996b). Using higher-taxon richness as a surrogate for species richness: II. Local applications. *Proc. Roy. Soc. London, Ser. B* **263**, 1571–1575.

Bryant, D., and Barber, C. (1994). In *World Resources: 1994–95,* pp. 147–164. World Resources Institute/Oxford University Press, New York.

Church, R. L., Stoms, D. M., and Davis, F. W. (1996). Reserve selection as a maximal covering location problem. *Biol. Conservation* **76**, 105–112.

DeLong, D. C. (1996). Defining biodiversity. *Wildl. Soc. Bull.* **24**, 738–749.

Dodson, S. I., Allen, T. F. H., Carpenter, S. R., Ives, A. R., Jeanne, R. L., Kitchell, J. F., Langston, N. E., and Turner, M. G. (1998). *Ecology.* Oxford University Press, New York.

Fiedler, P. L., and Jain, S. K. (eds.). (1992). *Conservation Biology: The Theory and Practice of Nature Conservation, Preservation and Management.* Chapman and Hall, New York.

Fisher, R. A., Corbet, A. S., and Williams, C. B. (1943). The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Anim. Ecol.* **12**, 42–58.

Gaston and Spicer (1998). *Biodiversity: An Introduction.* Blackwell Science, Oxford, United Kingdom.

Ghilarov, A. (1996). What does "biodiversity" mean—Scientific problem or convenient myth? *Trends Ecol. Evol.* **11**, 304–306.

Ghilarov, A., and Timonin, A. G. (1972). Relations between biomass and species diversity in marine and freshwater zooplankton communities. *Oikos* **23**, 190–196.

Haffer, J. (1969). Speciation in Amazon forest birds. *Science* **165**, 131–137.

Harper, J. L., and Hawksworth, D. L. (1994). Biodiversity: Measurement and estimation. *Philos. Trans. Roy. Soc. London, Ser. B* **345**, 5–12.

Hubbell, S. P. (1997). A unified theory of biogeography and relative species abundance and its application to tropical rain forests and coral reefs. *Coral Reefs* **16**, S9–S21.

Hudson, W. E. (ed.). (1991). *Landscape Linkages and Biodiversity.* Island Press, Washington, D.C.

Hunter, M. L., Jr. (1996). *Fundamentals of Conservation Biology.* Blackwell Science, Cambridge, Massachusetts.

Huston, M. A. (1994). *Biological Diversity: The Coexistence of Species on Changing Landscapes.* Cambridge University Press, New York.

IUCN/UNEP/WWF. (1980). *World Conservation Strategy: Living Resource Conservation for Sustainable Development.* International Union for Conservation of Nature and Natural Resources/U.N. Environment Programme/World Wildlife Fund, Gland, Switzerland.

Knopf, F. L. (1992). Focusing conservation of a diverse wildlife resource. *Trans. North Am. Wildl. and Nat. Resources Conf.* **57**, 241–242.

Lawton, J. H. (1994). Population dynamic principles. *Philos. Trans. Roy. Soc. London, Ser. B* **334**, 61–68.

Lovejoy, T. E. (1980). *The Global 2000 Report to the President* (G. O. Barney, ed.), Vol. 2, The Technical Report, pp. 327–332. Penguin, New York.

MacArthur, R. H. (1957). On the relative abundance of bird species. *Proc. Natl. Acad. Sci. USA* **45**, 293–295.

MacArthur, R. H., and MacArthur, J. (1961). On bird species diversity. *Ecology* **4**, 594–598.

MacArthur, R. H., and Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey.

Margalef, R. (1957). La teoria de la informacíon en ecologia. *Memorias de la Real Academia de Ciencias y Artes* (*Barcelona*), 3rd series **32**, 373–449.

McNeely, J. A., Miller, K. R., Reid, W. V., Mittermeier, R. A., and Werner, T. B. (1990). *Conserving the World's Biological Diversity*. International Union for Conservation of Nature and Natural Resources/World Resources Institute/Conservation International/World Wildlife Fund–U.S./World Bank, Gland, Switzerland/Washington, D.C.

Milner-Gulland, E. J., and Mace, R. (1998). *Conservation of Biological Resources*. Blackwell Science, Oxford, United Kingdom.

Mittermeier, R. A., Carr, J. L., Swingland, I. R., Werner, T. B., and Mast, R. B. (1992). Conservation of amphibians and reptiles. In *Herpetology: Current Research on Amphibians and Reptiles*. (K. Adler, ed.), pp. 59–80. Society for the Study of Amphibians and Reptiles, Contributions to Herpetology.

Myers, N. (1990). Threatened biotas: "Hotspots" in tropical forests. *Environmentalist* **8**, 1–20.

Noss, R. F. (1990). Indicators for monitoring biodiversity: A hierarchical approach. *Conservation Biol.* **4**, 355–364.

Noss, R. F., and Cooperrider, A. Y. (1994). *Saving Nature's Legacy: Protecting and Restoring Biodiversity*. Island Press, Washington, D.C.

Office of Technology Assessment. (1987). *Technologies to Maintain Biological Diversity*. *Summary*. U. S. Govt. Printing Office, Washington, D.C.

Pearson, D. L., and Cassola, F. (1992). World-wide species richness patterns of tiger beetles (Coleoptera: Cicindelidae): Indicator taxon for conservation and biodiversity studies. *Conservation Biol.* **6**, 376–391.

Pielou, E. C. (1977). *Mathematical Ecology*. John Wiley & Sons, New York.

Reid, W. V., and Miller, K. R. (1989). *Keeping Options Alive: The Scientific Basis for Conserving Biodiversity*. World Resources Institute, Washington, D.C.

Ricklefs, R. E., and Schluter, D. (1993). *Species Diversity in Ecological Communities*. University of Chicago Press, Chicago.

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge University Press, Cambridge, United Kingdom.

Samson, F. B., and Knopf, F. L. (1994). A framework to conserve biological diversity through sustainable land management. *Trans. North Am. Wildl. and Nat. Resources Conf.* **59**, 367–377.

Scott, J. M., Abies, E. D., Edwards, T. C., Jr., Eng, R. L., Gavin, T. A., Harris, L. D., Haufler, J. B., Healy, W. M., Knopf, F. L., Torgerson, O., and Weeks, H. P., Jr. (1995). Conservation of biological diversity: Perspectives and the future for the wildlife profession. TWS committee report. *Wildl. Soc. Bull.* **23**, 646–657.

Scott, J. M., Davis, F., Csuti, B., Noss, R., Butterfield, B., Grove, C., Anderson, S, H., Caicco, S., Derchia, F., Edwards, T. C., Jr., Ullman, J., and Wright, R. G. (1993). GAP analysis: A geographic approach to protection of biological diversity. *Wildl. Monogr.* **123**, 1–41.

Soukhanov, A. H., Ellis, K. E., Harris, D. R., DeVenne, P. B., and Severynse, M. (eds.). (1988). *Webster's II New Riverside University Dictionary*. 1563 pp. The Riverside Publ. Co., New York.

Sousa, W. P. (1979). Disturbance in marine intertidal boulder fields: The nonequilibrium maintenance of species diversity. *Ecology* **60**, 1225–1239.

Swingland, I. R. (1997). Global conservation and the sciences: People, policy and pennies. In *Conservation, Restoration and Management of Tortoises and Turtles* (J. Van Abbema, ed.), pp. 453–464. New York Turtle and Tortoise Society, New York.

Swingland, I. R. (1999). Commercialisation, structure and sustainability of biodiversity conservation. In *New Practices in Integrated Protected Area Management* (M. Walkey, I. R. Swingland, and S. Russell, eds.). Chapman and Hall, London.

Temple, S. A. (1986). Predicting impacts of habitat fragmentation on forest birds: A comparison of two models. In *Wildlife 2000: Modelling Habitat Relationships of Terrestrial Vertebrates* (J. Verner, M. L. Morrison, and C. J. Ralph, eds.), pp. 301–304. University of Wisconsin Press, Madison.

Thomas, R. (1992). Genetic diversity. In *Global Biodiversity: Status of the Earth's Living Resources:* (World Conservation Monitoring Centre, eds.), pp. 1–6. Chapman and Hall, London.

Vane-Wright, R. I. (1992). Systematics and diversity. In *Global Biodiversity: Status of the Earth's Living Resources:* (World Conservation Monitoring Centre, eds.) pp. 7–12. Chapman and Hall, London.

Vane-Wright, R. I. (1996). Systematics and the conservation of biological diversity. *Ann. Missouri Botanic Gardens.*

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* **1**, 213–251.

Williams, C. R. (1964). *Patterns in the Balance of Nature*. Academic Press, New York/London.

Williams, P. H., Humphries, C. J., and Vane-Wright, R. I. (1991). Measuring biodiversity: Taxonomic relatedness for conservation priorities. *Australian Systematic Botany* **4**, 665–679.

Wilson, E. O. (1961). The nature of the taxon cycle in the Melanesian ant fauna. *Am. Nat.* **95**, 169–193.

Wilson, E. O. (1992). *The Diversity of Life*. W. W. Norton & Co., New York.

Wilson, E. O., and Peter, F. M. (eds.). (1988). *Biodiversity*. National Academy Press, Washington, D.C.

Zonneveld, I. S. (1988). Composition and structure of vegetation. In *Vegetation Mapping* (A. W. Küchler and I. S. Zonneveld, eds.), pp. 25–35. Kluwer Academic Publisher, Dordrecht, Netherlands.

BIODIVERSITY, EVOLUTION AND

Gregg Hartvigsen
State University of New York, College at Geneseo

GLOSSARY

biodiversity A measure of the relative abundance of species or higher taxonomic units found in a certain area at a particular time. In ancient communities this is often simply referred to as the number of species or taxa found at a site, usually referred to as "species richness" in the ecological literature.

community A group of interacting organisms, representing multiple species in a given location.

complex adaptive systems A group of individuals or types that exhibit variability in which large-scale patterns emerge from small-scale interactions. The system is adaptive in the sense that it is subject to extrinsic selection that leads to a change in structure and/or dynamics over time.

evolution The morphological or genetic change in species over time. Small changes that do not lead to reproductive isolation among members of a group are referred to as "microevolution." Speciation, or the generation of new species, is generally referred to as "macroevolution."

fossil Anything found in strata of rocks or sediments that is recognized as the remains of an organism from a former geological time.

species A group composed of individuals that interbreed or potentially interbreed and are separated reproductively from other such groups.

taxon (taxa) A group of individuals representing a classification of organisms, such as genus, family, or order. Higher taxa are those above the species level.

EVOLUTION has led to an incredible diversification of forms that have originated from a single ancestor. Our current accounting of species on Earth is woefully lacking, as many of the entries in this encyclopedia have demonstrated. In addition, the fossil record, unfortunately, is grossly underrepresented by taxa that have existed on Earth since life began 3.5–4.0 billion years ago. The purpose of this article is to summarize our current understanding of the rise of Earth's biodiversity and to discuss how evolution has driven this process of the accumulation of taxa over time and how evolution influences communities, both past and present, in which organisms reside. The influence of evolution on integrated communities of diverse taxa and the likely changes that will take place in the near future certainly rank among the great frontiers facing students of biodiversity as we watch the profound test of the resilience of Earth's ecosystems to the enterprise of *Homo sapiens*.

I. INTRODUCTION

The earliest record of life on Earth is from approximately 3.6 billion years ago. Aside from a few difficult to test hypotheses about how life arose, a topic beyond the scope of this article, the fossil record clearly reveals that early life forms were single-celled organisms of the group Archaea with structures similar to modern bacteria. Earth at that time was far different than today and would have been uninhabitable to the vast majority of contemporary organisms. Our understanding of these life forms that have arisen over this time to the present relies mainly on the fossilized remains from organisms representing an array of groups that far exceeds those living today. Although the fossil record contains discontinuities and absences of taxa, recent evidence is sufficient to allow discussions of the evolution of communities, or groups of species, over time. This level of detail is astounding, but it is only suggestive of the myriad of information likely to be gained in the coming decades. In addition, recent development and use of computer models is enabling us to both understand these paleobiological data and make predictions of how evolution influences the dynamics of communities over time. I suggest that the coupling of our rapidly increasing knowledge of the fossil record and our use of computer modeling will help us to both understand and predict the influence of humans on patterns of biodiversity.

II. THE FOSSIL RECORD

A. What Does the Fossil Record Indicate?

In general, the fossil record is an outstanding record of changes in biodiversity, revealing both an increase and turnover in taxa over time. Data for marine organisms suggests that the diversity of families has generally increased during the Phanerozoic, which includes the past 540 million years, the time for which fossilized remains of multicellular organisms exist (Fig. 1; Sepkoski, 1991). A broader survey shows that diversity has increased exponentially for all known families, genera, and species during this period (Fig. 2; Benton, 1995). Increases in taxa, however, have been discontinuous, with major catastrophic losses of taxa in the form of mass extinctions, such as that which occurred at the end of the Cretaceous period that led to the loss of the dinosaurs. Thus, the accumulation of biodiversity in the form of numbers of taxa has been discontinuous and has exhibited a skewed distribution in the magnitude of the size of extinction events, ranging from relatively low-level background extinction rates to large-scale mass extinction events.

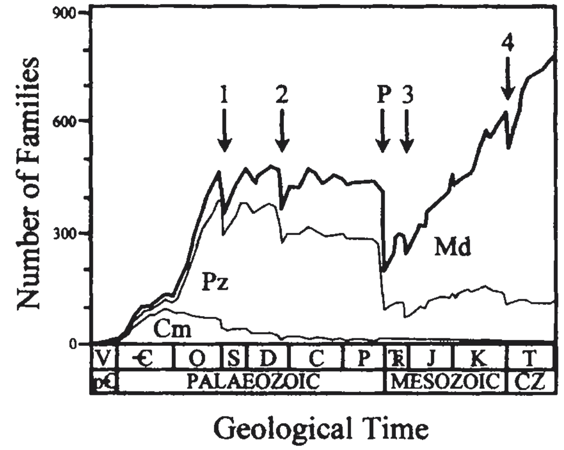


FIGURE 1 The diversity of skeletonized marine animal families during the Phanerozoic. The bold line represents total numbers of families known from the fossil record. Five major extinction events are indicated: 1, end Ordovician; 2, Late Devonian; P, end Permian (largest); 3, end Triassic; 4, end Cretaceous. V, Vendian; Ꞓ, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; TR, Triassic; J, Jurassic; K, Cretaceous; T, Tertiary. Faunal assemblages are categorized as Cambrian (Cm), Paleozoic (Pz), and Modern (Md). (after Sepkoski, 1991).

In general, we know that species are generally short-lived, averaging a duration on the order of 5 million years, with few taxa being long-lived. Most species are,

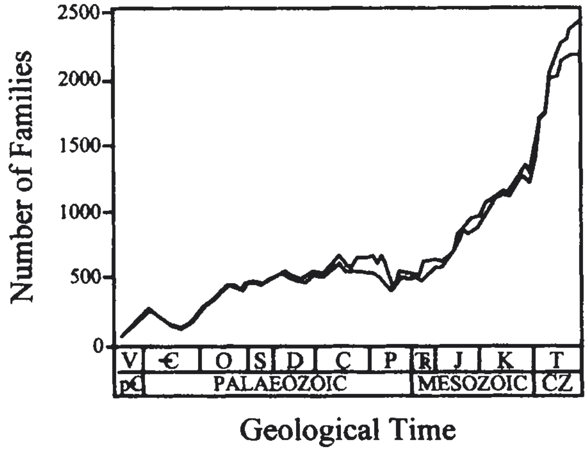


FIGURE 2 Diversity has increased exponentially for all known families, genera, and species during the Phanerozoic. The top and bottom lines represent maximum and minimum numbers of families, respectively. See the legend to Fig. 1 for abbreviations (after Benton, 1995).

and are inferred to have been, composed of relatively few individuals locally distributed with few species being relatively abundant and widely distributed. We also know that most genera have few species, whereas most families have few genera. These patterns have led and continue to lead to high probabilities that any particular taxon is likely to be relatively short-lived as opposed to long-lived, where the mean duration of taxa exceeds the median duration.

1. The Number of Taxa Has Increased over Time

A clear and dramatic trend exists in the fossil record of an increase in diversity. This pattern exists for taxa ranging from the level of species to phyla, and it has been well documented for plants, invertebrates, and vertebrates. This pattern is seen, for example, in the number of marine families over time during the Phanerozoic eon, since about 540 million years ago (Fig. 1). The major additions to diversity in the fossil record include skeletonized marine invertebrates during the Cambrian, especially trilobites, brachiopods, and archaeocyathans. Corals, cephalopods, ostracods, crinoids, and star fish arose through the remainder of the Paleozoic, and bivalves, gastropods, echinoids, teleost fish, and marine reptiles arose during the Mesozoic. Diversity increased on land and included the evolution of vascular plants (Silurian and Devonian), gymnosperms (Carboniferous), and angiosperms (Jurassic). Insect diversity exploded in the late Paleozoic, followed by social insects in the Jurassic. Major clades of terrestrial vertebrates arose by filling vacant niches, including amphibians (Devonian), reptiles (Carboniferous), pterosaurs and dinosaurs (Triassic and Jurassic), birds (Jurassic), and mammals in the early Tertiary.

The pattern, however, is one of fluctuations over time in the shorter term. The number of species of relatively large herbivores and carnivores in North America during the past 44 million years, for example, appears to exhibit little long-term trend toward increased or decreased diversity (Fig. 3). Interestingly, however, these data suggest that the number of predators depends on the number of potential prey species over time. There is a significant correlation between the number of predators and prey which is consistent with the hypothesis that interspecific associations are interdependent within communities over time.

2. Extinction Events Are Followed by Evolutionary Radiations

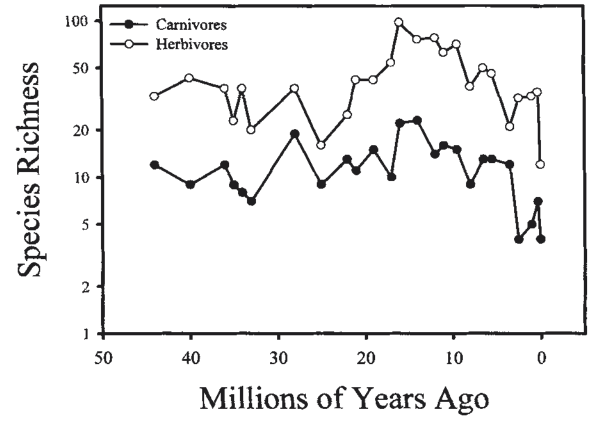


FIGURE 3 The number of large mammal species from North America during the past 44 million years. Note the correlation between the diversity of predators and their potential prey ($r^2 = 0.43$; $p < 0.001$) (data from Van Valkenburgh and Janis, 1993).

Extinction and the origin of species defines the dynamics of the fossil record. Species, for instance, typically last on the order of 1–10 million years and appear to go extinct in a non-random fashion. Therefore, species arise and ultimately disappear throughout time, producing a general pattern of loss referred to as "background" extinction (Fig. 4).

Biodiversity at various levels of organization has increased during the Phanerozoic. There exist, however,

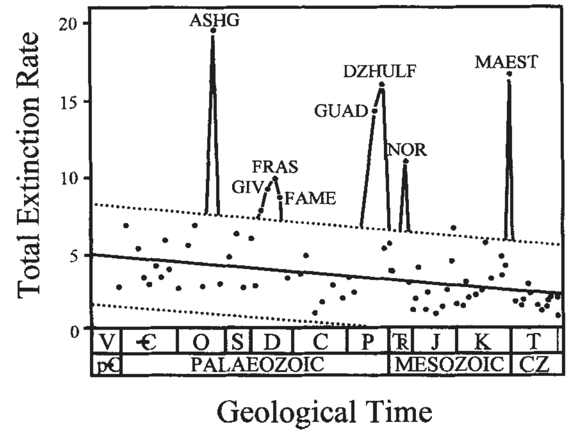


FIGURE 4 Background and mass extinction of marine animal families during the Phanerozoic. Rates are reported as the number of families lost per million years. Extinctions within dotted lines represent background rates, whereas solid lines above the upper dotted line represent mass extinctions. Mass extinctions: ASHG, Ashgillian (end Ordovician); GIV, Givetian; FRAS, Frasnian; FAME, Famennian (Late Devonian); GUAD, Guadalupian; DZHULF, Dzhulfian (end Permian); NOR, Norian (end Triassic); MAEST, Maastrichtian (end Cretaceous). For other abbreviations, see the legend to Fig. 1 (after Raup and Sepkoski, 1982).

at least five catastrophic extinction events that have led to a decrease in the number of taxa, and therefore biodiversity, that greatly exceed the background extinction rate (Fig. 4). In particular, the extinction event at the end of the Permian (Fig. 1, "P") resulted in a loss of a maximum estimated 96% of species on Earth (although this is likely an overestimate; Raup, 1994).

The overall pattern of extinction appears to follow a well-defined "kill curve," proposed by David Raup, that indicates a skewed distribution of species loss ranging over the gradient where most time periods exhibit low rates of extinction (background) and few periods of very high rates of extinction (mass extinctions). These data, therefore, suggest a predictable pattern, based on probability, that describes species extinctions.

Following extinctions there is general radiation, or rapid evolutionary diversification, of taxa. This is clearly seen in Sepkoski's database of marine families (Fig. 1) and even more clearly in Benton's larger database of families from both terrestrial and marine systems following the five major extinction events (Fig. 2).

The rates of species origination have varied over time and are highest following periods of mass extinction and in times when new ecospace became available, such as during the initial rise of diversity during the late Vendian and Cambrian (Figs. 1 and 2). Speciation was highest for terrestrial families during the Priabonian (late Eocene), most likely due to the increase in insect diversity.

B. How Good Is the Fossil Record?

The fossil record is far from a complete representation of all of Earth's biotic history. There are only bits and pieces of evidence from organisms and we are able to date strata from which organisms derive only with varying levels of accuracy. The fossil record is better, in general, for recent strata and decreases with age. The fossils we do have, however, enable us to construct a logical progression in both complexity of form and diversity over time. The idea that evolution is progressive, however, has long been abandoned in light of the many examples of lineages that do not follow straight progressions of simple to more complicated forms over time. Moreover, extant species that appear ancient, such as the coelacanth, horseshoe crab, and cockroach, are neither unchanged from their ancestors nor is their longevity unexpected. By chance we expect variance in the longevity among lineages, with many short-lived species and few long-lived species.

Understandably, only certain morphological characteristics have been susceptible to fossilization. Of those that are fossilized we often find only disarticulated parts of organisms, such as a tooth, part of a shell, or perhaps a bone fragment, separated over varying distances. Soft-bodied organisms and soft tissues are rarely observed. To complicate matters, these fragments are susceptible to movement both spatially (e.g., down-stream after death) and vertically in the stratigraphy, which makes fossils appear to move temporally. This occurs when sedimentation rates are very low, allowing noncontemporaneous organisms to become fossilized together, or when layers are mixed, a process referred to as "time averaging" of the fossil record. However, taphonomy, or the study of postmortem behavior of organic remains, indicates that a vast majority of organisms are not transported out of their original habitats and therefore can be used to construct probable past communities. Evidence for this derives primarily from the observation of contemporary local death assemblages in which individuals rarely move across community boundaries.

Microfossils such as pollen are abundant in lake sediment and can be used for quantitative reconstruction of local plant communities. Pollen from sediments on the order of a few hundred thousand years has been used to reasonably describe assemblages. Little is known about the role that evolution played in plant communities from sediments more than 1 million years old. The pollen record during the Holocene (e.g., since the last glacial period ended about 16,000 years ago) has been widely and more confidently used to reconstruct past plant communities. Therefore, we remain perplexed with regard to the various assemblages of taxa that likely resided concurrently in more ancient communities, but for more recent times we have increasing confidence in our effort to reconstruct communities from fossil remains.

C. What Does the Fossil Record Indicate About Communities?

The interpretation of community structure in the fossil record is more problematic than documenting the change in taxa over time. The primary difficulty lies in determining whether fossilized taxa occurred contemporaneously. Postmortem movement of organisms and their parts appears to take place over relatively short distances and within time-averaged ranges of species, which is a pattern found for both marine and terrestrial taxa. Therefore, optimistically, it appears that co-occurring taxa through the fossil record likely represent evidence of past communities.

Evidence from pollen records, however, can produce a paradox in our interpretation of past plant communities. Over long periods of time pollen deposition in lake sediments generally accurately reflects nearby plant communities. The distance-dependent deposition of pollen is generally correlated with the sum basal area of trees (cross-sectional area of tree diameters) within a few square kilometers of the surrounding area, although many factors influence the pollen collected from lakes and the surrounding vegetation, such as relative decomposition rates and the relative efficiency of a lake to capture pollen. Pollen records found in lake sediments, however, suggest that plant community change, in response to rapid climate change following the Wisconsin glacial period in North America, is the product of independently acting species. In addition, there is a consistent pattern that, among late Quaternary pollen records for basins of different sizes and from different regions (temperate and tropical), communities appear to be the product of independently acting species.

The rain of leaf litter and other debris, likewise, is well correlated in contemporary communities with the abundance of trees within local communities. Unfortunately, leaf litter in general rapidly decomposes and rarely provides evidence of past communities.

The fossil record remains difficult to interpret with regards to community composition and dynamics over time. Much research is needed and is under way on the dynamics of communities over long periods of time. In particular, we need to resolve conditions that lead to the following paradox: Taphonomy of current assemblages suggests fossil taxa may represent contemporaneous communities, whereas pollen studies suggest that at least plant species have acted relatively independently over time. Additional data are needed to help accurately reconstruct the composition of past communities.

III. CURRENT PATTERNS OF BIODIVERSITY

Biodiversity is heterogeneously distributed on Earth, with the general pattern that species densities are greatest near the equator and decrease with increasing latitude. Hypotheses and evidence for these patterns abound. In the following sections, I briefly introduce how we define biodiversity, discuss very general trends in communities, and apply these to understanding biodiversity of communities over time.

A. How Do We Define Biodiversity?

Before discussing patterns of biodiversity change over time, I need to define diversity. In general, paleontologists consider biodiversity to be equivalent to the number of unique taxa. In the ecological literature, diversity generally incorporates both the number of species and equitability or the relative abundance of individuals among species. A simple hypothetical example is the relative diversities of two communities. Assume that both have 100 individuals within two species. One community has 99 individuals of one species and the second community has 50 individuals of both species. The first community is essentially a monoculture of species 1. The second community is considered a more diverse community, despite the fact that both communities have the same number of species present.

Accounting for equitability in the fossil record, however, is quite problematic because of the spotty nature of the preserved specimens. It must be recognized that fossils in an assemblage will not likely constitute a random sample of individuals from a community. Rather, the remains of fossil communities are likely composed of a skewed distribution of individuals representing only a subset of those species that contained structures susceptible to fossilization. Other species likely escaped fossilization completely or were missed in the sample from which the fossils were found. In addition, the subset that is found is more than likely to be small and therefore highly susceptible to the variance incurred from the chance occurrence of individuals of rare species. This is an example of the problem referred to as the "central limit theorem" in statistics in which sampling of a skewed distribution (the reality of the fossil communities) yields a normally distributed population (not an accurate representation of the fossil communities). In other words, we are most likely to find fossils of the most abundant species, yielding an underestimate of the true number of species that occurred at that time in that community. Some fossil assemblages, however, will contain samples of relatively rare taxa, making these taxa appear more abundant that they really were. In addition, it is difficult to assess whether forms represent truly different taxa. We are more likely than not to lump taxa together because of our inability to know whether similar fossils represent members of an interbreeding population.

B. Patterns of Species Change Over Time

Simple change over time can be expressed as a rate, similar to the change in a population over time. We can represent this change in the number of taxa over time as $dN/dt = RN$, where N represents the number of species in a genus or genera in a family, and $R = S - E$, where S is the speciation rate and E is the extinction rate. We may therefore assume $dN/dt = 0$ to represent a null hypothesis against which we test for significant changes in the rates of species additions or subtractions. There are several problems with this approach, however, since our null hypothesis is that all species are expected to eventually become extinct. Therefore, we must define a time constraint of interest, concluding that if $R < 0$ then the taxa of interest appear to be on the decline. Comparisons of the fossil record have been found to differ from this random null hypothesis. The database assembled by Sepkoski (Fig. 1) suggests that diversity increases logistically, whereas the data set assembled by Benton (Fig. 2) suggests diversity has increased exponentially during the past 600 million years.

C. Are Communities Saturated with Species?

In general, it appears that many communities are at or near a saturation level for the number of species they contain. The evidence for this derives indirectly from several sources, including the relationship of families depicted previously in the fossil record and from the ecological literature in which there is found a general inability of introduced species to successfully invade communities. Species introductions usually fail, and when they succeed species usually have little consequence on natural systems, despite the numerous celebrated examples of introduced species altering native communities. Occasionally, even with the loss of native dominant species such as the American chestnut tree to a fungal blight, the effect on natural processes may be equivocal.

The result of this contemporary observation suggests that communities may reach taxa saturation. If this pattern holds for prior paleocommunities, then this suggests that ecological and evolutionary processes operate to fill available niche space with either colonizing species or through the process of speciation. Inspection of either Sepkoski's or Benton's databases for the number of taxa over time suggests that the Earth is not at a carrying capacity. The following is the obvious question stemming from this: Where should we expect to find new species occurring, especially in light of the estimation that humans are currently consuming on the order of 40% of the Earth's productivity?

IV. COMMUNITY STABILITY

Researchers in the field of ecology, and particularly conservation biology, have long been interested in the dynamics of communities and the rate of species change over time, also referred to as "species turnover." George Evelyn Hutchinson and his student Robert MacArthur were early advocates of the notion that stability of communities is positively related to the number of species residing in the community. This relationship, quantified by MacArthur and incorporated into the influential theory of island biogeography developed in collaboration with E. O. Wilson, portrayed ecological systems as having stable equilibria, resulting from the trade-off between colonization and extinction rates on islands. Recently, this theory has been extended to systems that, in general, function like islands.

Although intuitively attractive, the relation between diversity and stability may not exist, at least in some systems. In 1973, Sir Robert May polarized the field of community ecology by introducing a simple matrix model approach that suggested an inverse relationship between species diversity and community stability. Recently, empirical studies have suggested that diversity and stability are related. Research by David Tilman on grasslands in Minnesota, for instance, suggests that community resistance to drought is correlated with community diversity.

A. Communities as Complex Adaptive Systems

Recent developments in our understanding of community assembly, organization, and development over ecological and evolutionary time suggest that communities are more complex than previously believed. Classic approaches to understanding the dynamics of communities portrayed them as being composed of individual species that interact ecologically with little regard to spatial location or temporal dynamics. Recently, we have come to appreciate the intricacies of small-scale spatial and temporal partitioning of resources and interactions among co-occurring organisms. In addition, in the past 30 years we have come to appreciate the immense genetic diversity within species that enables species to adapt to biotic and abiotic factors that operate spatially within communities (Thompson, 1994).

Complex adaptive systems can be defined as groups of individuals that exhibit variability, interact with a subset of individuals within the global population (usually within some local neighborhood), and are susceptible to an autonomous, selective process that usually leads to individuals experiencing differential reproduction. These rules apply quite generally to species, communities, and other systems such as economies, allowing them to adjust to changing environmental conditions. Natural communities fit this general class of systems very well, and recent theoretical work has described the similar dynamics among diverse systems. Recent work, for instance, suggests that population and community dynamics may be influenced by the process of evolution acting on genetic diversity among individuals (Hartvigsen and Levin, 1997). It is likely that this new theoretical approach will lead to a greater understanding of how communities respond evolutionarily to disturbances over time scales greater than years.

B. The Effect of Disturbance on Community Biodiversity: Evolutionary versus Ecological Time

The causes of various mass extinction events are not known, but hypotheses do exist. The most recent mass extinction at the end of the Cretaceous was most likely caused by an asteroid impacting the Earth, forming the Chicxulub crater on the north side of the Yucatan Peninsula. The causes of the other mass extinctions are unresolved but may be the result of similar massive disturbances. The probability distribution of taxa loss appears to follow what we expect to occur by chance and matches the loss of taxa due to random factors. Occasionally, we should expect to lose many species, whereas over most time periods only a few species should be expected to be lost. Therefore, the greater the disturbance, the greater the effect (reduction) on biodiversity.

The probability of species loss depends on many factors, including the size of the populations of species within the community. Smaller populations are more likely to go extinct than larger populations, a process referred to as demographic stochasticity. The Gambler's Ruin is a metaphor that is commonly associated with this process. The idea is that a gambler in a casino who has a limited number of betting units to lose (e.g., quarters for a slot machine or equivalents to the minimum bet at a black jack table) is more likely to go bankrupt simply by chance compared to a gambler with many betting units. Likewise, stochastic, or random, changes in population size lead to small populations being more likely to reach the absorbing state of zero individuals than a large population. Therefore, and not surprisingly, small populations are at greatest risk of extinction simply by chance.

Communities composed of species with small populations are therefore likely to suffer higher levels of species loss than communities composed of species with large populations. Raup (1994) argues that there appears to be little predictability in which species will be eliminated and that extinction does not appear to be selective. Species with narrow geographic ranges may be more susceptible to extinction, whereas species with large geographic ranges are not likely to experience disturbances so great as to influence their entire population.

Small populations typically exhibit lower genetic variability than large populations and are more likely to become locally extinct. Small populations, however, may experience more rapid evolution due to reduced genetic inertia. This may be seen in the founder effect, in which a few individuals begin a new population leading to allopatric speciation.

Without knowing the probability distribution of extrinsic factors causing large-scale mass extinctions, it is difficult to extrapolate from the fossil record if species that went extinct in fact had small populations. In addition, it does not appear that evolution during the past several hundred years has reduced the likelihood of extinction events. There is no clear reduction in the apparent risk of such mass extinctions with time in the fossil record. Using the "all things being equal" assumption, the fossil record (Figs. 1 and 2) suggests that, with a general trend toward increased diversity over time, the condition of having more taxa should lead to increases in the number of taxa lost due to disturbances. That is, if the probability of extinction of any particular species is constant, then we should see increased extinction rates as diversity has increased. However, as shown in Fig. 4, the background extinction rate appears to have decreased with time. There remains the need to evaluate the risk of extinction by determining factors such as the absolute abundance of individuals within species and how this influences extinction risk. For example, if increased diversity over time has led to smaller population sizes due to ecological partitioning of resources, we may anticipate a trend toward increased extinction rates.

Experimental work on laboratory and field communities suggests that decreasing biodiversity also leads to a reduction in ecological time of ecosystem function, such as total productivity and carbon dioxide sequestration (Tilman and Downing, 1994; Naeem *et al.*, 1994). This suggests that a positive feedback mechanism may operate so that disturbed communities are more likely to lose species in addition to those initially lost due to decreased population size.

C. Models of the Influence of Evolution on Ecological Systems

The effect of evolution on the likelihood of species extinction and community stability represents a fundamental problem in the field of evolutionary ecology and is of great importance to our ability to conserve species over long time periods. Theoretical work suggests that the process of evolution influences population and community dynamics (Hartvigsen and Levin, 1997; Doebeli, 1996). Hartvigsen and Levin, using a spatially explicit, individual-based model that accounts for genetic variability at the individual level, leading to evolution over time, show that the process of evolution in a model system can slow population dynamics, suggesting that the risk of species extinction declines due to reduced fluctuations resulting from demographic stochasticity. If this occurs in real systems, then we might expect a positive feedback mechanism to operate such that a population reduced in size is likely to experience a loss in genetic diversity. In addition to other known factors that increase extinction risk of small populations, including inbreeding depression and social dysfunction (Allee effect), a loss in genetic diversity may increase population fluctuations, thus increasing the likelihood that such a species will go extinct.

V. THE FUTURE RECORD

A. Documenting Earth's Current Mass Extinction Event

The extent of the current extinction event will only be known in retrospect. However, there are patterns that are recognizable today. For example, there exists a good record of species that have gone extinct during the past 400 years (Smith *et al.*, 1993). A minimum of 485 animal species and 584 plant species became extinct during this period, with approximately half of these occurring during the twentieth century. This includes the loss of nearly 1% of all known bird species during the past 100 years, mostly due to human activity. This loss suggests an average species duration of about 10,000 years, which is two or three orders of magnitude shorter than has occurred on average in the fossil record (in other words, this represents an extinction rate two or three orders of magnitude greater than the background extinction rate in the fossil record). It is this rate of known extinction, compared to average extinctions in the fossil record, that suggests humans are the impetus for a mass extinction event.

B. The Big Question: How Will the Process of Evolution Influence Biodiversity Dynamics?

Evolution generates new species in the aftermath of mass extinction. This is well supported by the fossil record and appears, in geological time, to be rapid. The current extinction event taking place on Earth, however, appears to be more rapid than has occurred previously. Speciation will undoubtedly lead to the rise of new taxa but over time scales that are likely to be too long to have a noticeable effect in our lifetime.

The other role that evolution plays in biodiversity is its ability to influence the stability of communities. Little is known about this role of evolution. Empirical data for plant communities suggest that stability is correlated with biodiversity, such that more diverse communities are more resilient to disturbance. Natural communities are often composed of genetically variable individuals, which provide the foundation that enables species to evolve in response to biotic and abiotic factors over time. Also, communities with more species are in general more diverse genetically. If the time frame of environmental change can be matched by selection operating on genetic diversity, communities may resist disturbances by adjusting to changes over time. The analogy of species wandering over an "adaptive landscape" (sensu Sewall Wright) produces an image of communities functioning as complex adaptive systems. Ecosystems that function as complex adaptive systems may resist environmental fluctuations. If, however, environmental disturbances are abrupt, communities may either not be able to react quickly enough or suffer overloads that essentially shake species out of the community. There is evidence that communities harbor redundancy at the species level, but we have not yet determined patterns that identify which species are necessarily important or unimportant to community function. Therefore, the loss of species from communities is likely to decrease, if only slowly, the stability of communities.

Theoretical work suggests that the process of evolution may increase the resilience of communities to disturbance. This area of inquiry is in its infancy and will likely become an important and rapidly developed discipline in the following decades.

VI. CONCLUSION

Biodiversity, measured as the number of taxa, has increased on Earth over time. This pattern appears to follow either an exponential or a logistic growth pattern of taxa during the past 540 million years. The increase in taxa, however, has been discontinuous due to extinction events. There appears to be a probability distribution of extinctions with small background extinctions occurring frequently and large, catastrophic mass extinctions occurring rarely. Biodiversity has repeatedly and rapidly rebounded following such extinction events.

Biodiversity currently appears to be near a saturation level, based on the observation that most species that are introduced to areas fail to successfully become established. However, the pattern of increase in the number of taxa (e.g., families) over time does not appear to be at or near an asymptote (Fig. 2). The extreme alteration of the earth's biosphere by humans is likely to lead to a mass extinction event, currently on the order of 10–100 times the background extinction rate observed in the fossil record. In areas in which habitats are disturbed rather than destroyed, an explosive radiation of new species may occur, manifested, however, only over geological time.

The process of evolution may help stabilize communities. However, current human-induced worldwide disturbances are likely to lead to reduced population sizes and a resulting loss in genetic diversity, ultimately increasing the likelihood that more species will go extinct.

See Also the Following Articles

ARCHAEA, ORIGIN OF • DIVERSITY, COMMUNITY/REGIONAL LEVEL • EXTINCTION, RATES OF • FOSSIL RECORD • LIMITS TO BIODIVERSITY (SPECIES PACKING) • MASS EXTINCTIONS, NOTABLE EXAMPLES OF • NORTH AMERICA, PATTERNS OF BIODIVERSITY IN

Bibliography

Bambach, R. K., and Bennington, J. B. (1996) Do communities evolve? A major question in evolutionary paleoecology. In *Evolutionary Paleobiology* (D. Jablonski, D. H. Erwin, and J. H. Lipps, Eds.), pp. 123–160. Univ. of Chicago Press, Chicago.

Benton, M. J. (1995). Diversification and extinction in the history of life. *Science* **268**, 52–58.

Davis, M. G. (1981). Quaternary history and the stability of forest communities. In *Forest Succession: Concepts and Applications* (D. C. West, H. H. Shugart, and D. B. Botkin, Eds.), pp. 132–153. Springer-Verlag, New York.

Delcourt, H. R., and Delcourt, P. A. (1991). *Quaternary Ecology: A Paleoecological Perspective*. Chapman & Hall, New York.

Doebeli, M. (1996). Quantitative genetics and population dynamics. *Evolution* **50**, 532–546.

Flessa, K. W., and Kidwell, S. M. (1995). The quality of the fossil record: Populations, species, and communities. *Annu. Rev. Ecol. Syst.* **26**, 269–299.

Hartvigsen, G., and Levin, S. (1997). Evolution and spatial structure interact to influence plant–herbivore population and community dynamics. *Proc. R. Soc. London Ser. B* **264**, 1677–1685.

Jablonski, D. (1999). The future of the fossil record. *Science* **284**, 2114–2116.

Levin, S. A. (1999). *Fragile Dominion*. Perseus, Reading, MA.

Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H., and Woodfin, R. M. (1994). Declining biodiversity can alter the performance of ecosystems. *Nature* **368**, 734–737.

Raup, D. M. (1991). *Extinction: Bad Genes or Bad Luck*? Norton, New York.

Raup, D. M. (1994). The role of extinction in evolution. *Proc. Natl. Acad. Sci.* **91**, 6758–6763.

Raup, D. M., and Sepkoski, J. J., Jr. (1982). Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503.

Sepkoski, J. J., Jr. (1991). Diversity in the Phanerozoic oceans: A partisan review. In *The Unity of Evolutionary Biology* (E. C. Dudley, Ed.), pp. 210–236. Dioscorides Press, Portland, OR.

Smith, F. D. M., May, R. M., Pellew, R., Johnson, T. H., and Walker, K. S. (1993). Estimating extinction rates. *Nature* **364**, 494–496.

Thompson, J. N. (1994). *The Coevolutionary Process*. Univ. of Chicago Press, Chicago.

Tilman, D., and Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature* **367**, 363–365.

Van Valkenburgh, B., and Janis, C. M. (1993). Historical diversity patterns in North American large herbivores and carnivores. In *Species Diversity in Ecological Communities: Historical and Geographical Perspectives* (R. Ricklefs and D. Schluter, Eds.), pp. 330–340. Univ. of Chicago Press, Chicago.

BIODIVERSITY GENERATION, OVERVIEW

Paul H. Harvey
University of Oxford

GLOSSARY

key innovation A property of a species that promotes the process of speciation.

molecular phylogenies Phylogenies constructed from comparisons of homologous genetic material.

radiation All the species descended from a single common ancestor; tends to be a speciose group in comparison with others of a similar age.

reconstructed evolutionary trees Phylogenies linking contemporary species, therefore containing no information about lineages that have gone extinct.

sister taxa The expression has two uses, according to context. First, from a species perspective, the sister taxon of a particular species is the one with which it had the most recent common ancestor. Second, from the perspective of a node in a bifurcating phylogenetic tree, each daughter lineage gives rise to a set of contemporary species. The two sets are sister taxa with reference to the node.

BIODIVERSITY INVOLVES AT LEAST TWO COMPONENTS: (1) phylogenetic biodiversity, which might be summarized crudely as the number of species, and (2) character state biodiversity. To understand the generation of biodiversity one must understand both components. A description of the generation of phylogenetic biodiversity can only be properly achieved if there is a good fossil record. In the absence of a fossil record, molecular phylogenies can often be reconstructed to show when pairs of species last shared common ancestors in relative or absolute time. It is possible to analyze the structure of such phylogenies to estimate historical speciation and extinction rates, and to detect changes in speciation and extinction rates. The causes of variation in rates of speciation are becoming better understood through the analysis of possible key innovations that allow taxa to exploit more niches, with the associated methodology of sister taxon analysis. Reasons for character state diversity, such as variation in morphology, behavior, or life history, can be determined by using modern comparative methods. Since species inherit characters from their ancestors, closely related species are more similar than distantly related species, and species do not provide independent points for sta-

Encyclopedia of Biodiversity, Volume 1

tistical analysis. Instead, independent evolutionary origins of character states are sought. For continuously varying characters, such as body size, independent contrast analysis often provides a useful approach. The results of such analyses reveal that there are far fewer causal factors than character states, so that the same character complexes have evolved on many occasions, thereby simplifying a complex database. The consequence is that understanding the reasons for the generation of biodiversity becomes a manageable task.

I. THE FOSSIL RECORD AND CONTEMPORARY BIODIVERSITY

The generation of biodiversity involves, in large part, the origins of species. The fossil record, though notoriously incomplete, provides the most direct evidence for tracing temporal changes in biodiversity, including both the origin and extinction of lineages. Biases in preservation, often confounded with differences in taxonomic interest among paleontologists, mean that much more is known about the history of some taxa than others. For example, May (1996) points out that while around 95% of fossil databases consist of shallow-water marine invertebrates, most of today's known species are terrestrial, with 56% being insects. At the same time, Benton and Storrs' (1996) analyses suggest that more is known about the fossil history of vertebrates, most of which have been terrestrial, than of echinoderms, which are marine.

Benton's (1993) extensive compilations of stratigraphic incidence records for different lineages reveal the extent of well-known mass extinctions, followed by the subsequent regeneration of biodiversity. The rises in biodiversity that follow mass extinctions are presumably a consequence of the exploitation of (and speciation into) vacant niches. Also it is clear that the demise of some radiations is accompanied or followed by the rise of others; dinosaurs and mammals are cases in point. The causes of such complementarities have been hotly debated for many years. Similarly, there are times when one radiation is accompanied by, or shortly follows, another. The angiosperm plants are a massive radiation, as are the beetles. Cause and effect have been firmly suggested: the radiation of angiosperm plants provided a new set of niches into which beetles could speciate (Farrell, 1998). Arguments such as this are frequently anecdotal in that they involve post hoc explanations that require further testing, which often proves very difficult. For example, while the angiosperm plant radiation can explain about one-half of the species number in the beetle radiation (there are about 300,000 known beetle species), the other half remains unexplained. As Barraclough *et al.* (1998) point out, the angiosperm radiation does not explain the large radiation of 50,000 mainly carnivorous and fungivorous beetle species in the Cucujoidea or the equally large radiation of predatory beetle species in the Staphylinoidea.

II. RECONSTRUCTED VERSUS REAL EVOLUTIONARY TREES

The reasons for the origins of contemporary biodiversity are becoming better understood, largely as a consequence of advances in molecular genetics. Gene sequence analysis has become routine in many laboratories, and sequence data are deposited in openly accessible data banks. When equivalent sequences are available from different species, phylogenetic trees can be reconstructed. If a molecular clock has been operating, these trees show when in relative time any pair of species last shared a common ancestor. When the molecular clock can be dated, the trees can be calibrated in real time. Such phylogenies show the relationships among contemporary species, and contain no explicit information about extinct lineages. However, the structure of such trees can be used to demonstrate which evolutionary models are more likely than others to explain the origins of biodiversity in particular taxa.

When a model-based approach is used to analyze the origins of biodiversity, it soon becomes apparent how easily intuition can lead us astray. For example, the tree reproduced in Fig. 1 shows the relationships among contemporary salamander species from the genus *Plethodon*. The branching rate in the tree seems to have increased recently. This, however, does not necessarily mean that the net rate of speciation has increased. The simplest model to describe the generation of biodiversity is a constant-rates model in which the rate of speciation (lineage splitting) and the rate of extinction have been the same in all lineages at all times since the root of the tree. Obviously the rate of speciation will have been greater than the rate of extinction, or the tree would not have grown. Under such a model, if the rate of extinction is zero, a lineages-through-time plot based on the relationships among contemporary species would be a straight line with a slope, λ, the speciation rate. As the rate of extinction (μ) increases, so does curvature of the line upward toward the present (Fig. 2). This means that the pletho-

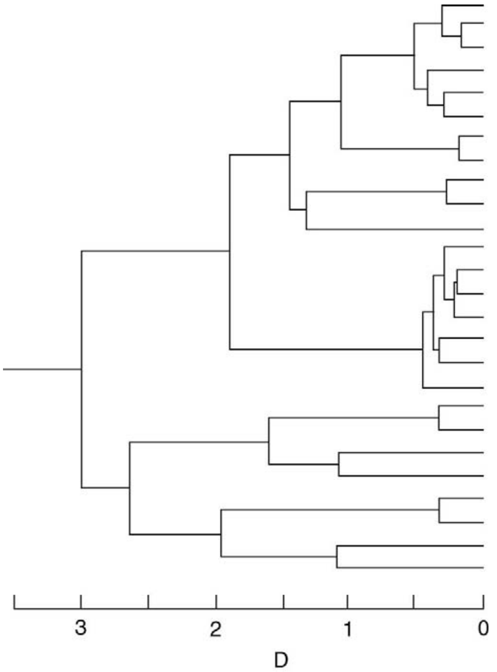


FIGURE 1 A phylogenetic tree reconstructed from genetic data showing the relationships among salamanders of the genus *Plethodon*. *D* is Rogers' genetic distance. The phylogeny reconstruction assumes a molecular clock, meaning that nodes are placed approximately correctly in relative time. No absolute timescale is available because the rate of ticking of the molecular clock is not known with precision. (After Nee, S., Holmes, E. C., May, R. M., and Harvey, P. H. (1995). Estimating extinction from molecular phylogenies. In *Estimating Extinction Rates* (J. H. Lawton and R. M. May, eds.) Oxford University Press, Oxford, United Kingdom, pp. 164–182.)

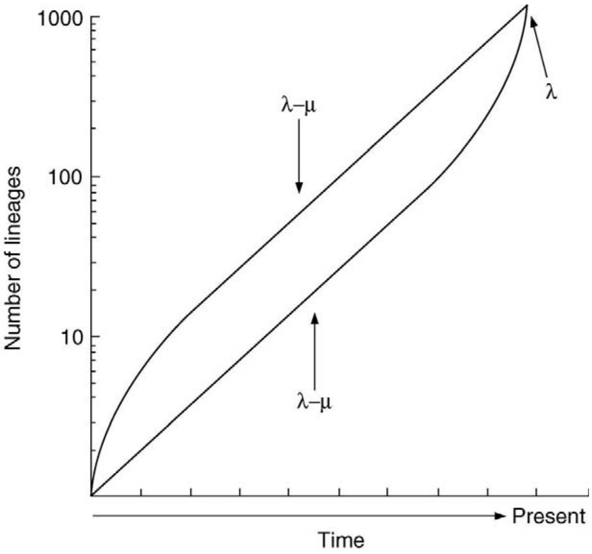


FIGURE 2 The expected forms of an actual phylogeny (top line) and a reconstructed phylogeny (bottom line) produced by a birth–death process, in which the constant per-lineage speciation rate (λ) is greater than the constant per-lineage extinction rate (μ). Note that the two curves would be superimposed if the extinction rate was zero. The initial steep slope of the line for the actual phylogeny is a consequence of only a sample of phylogenies surviving to the present and those are likely to be the ones that got off to a flying start. Note also that as μ increases from zero toward λ, so does the change in slope of the reconstructed line, so that a steep upturn describes a phylogeny where μ is high in comparison with λ. (After Harvey, P. H., May, R. M., and Nee, S. (1994). Phylogenies without fossils. *Evolution* 48, 523–529.)

dontid salamander tree, with its increased branching rate toward the present, is representative of a phylogeny generated under a high rate of extinction relative to speciation. Of course, speciation and extinction rates could have been changing through time, but the simple constant-rates model is sufficient to explain the data. If we accept this model, then it is also possible to estimate speciation and extinction rates or, more usefully for conservation biologists, the net rate of speciation ($\lambda - \mu$) and the danger of a clade going extinct (μ/λ) (Fig. 3). Similar analyses have been carried out for diverse taxa, including primates, carnivores, *Drosophila,* and birds.

Biodiversity cannot increase indefinitely. Eventually, any adaptive radiation will run out of niches to occupy and net speciation rates will decrease. Such an effect is apparent when Sibley and Ahlquist's (1990) molecular phylogeny of birds is analyzed as a lineages-through-time plot (Fig. 4). Instead of the line steepening toward the present, there is a general leveling off. In the absence of fossil material, which is relatively sparse for birds, it is not possible to say whether, in fact, speciation rates have decreased through time, extinction rates have increased, or both. Sibley and Ahlquist's phylogeny is one of the earlier molecular phylogenies and is based on DNA–DNA hybridization data: DNA from different species is annealed, and then heated to dissociate the strands. The more similar that strands are in base composition, the higher the temperature at which they dissociate. More recent studies based on direct comparison of sequences largely support Sibley and Ahlquist's original phylogeny, which, for example, indicated for the first time that the Australian passerines are an independent radiation from other passerines. In addition to many sequence-based studies that lend support for the general phylogeny, one reports a slowing down in the net rate of speciation in eight out of nine genera studied.

If we are to better understand the reasons for the generation of biodiversity, it is critical that molecular phylogenies are properly related to the known fossil record and to our understanding of biogeographic events. For example, the molecular phylogeny of birds

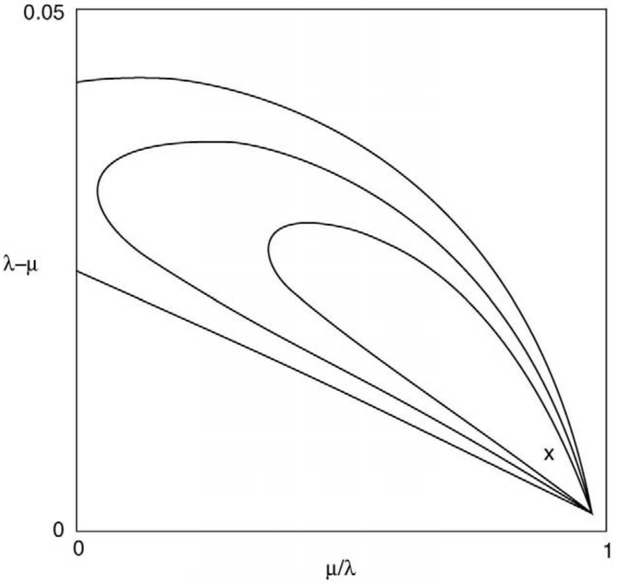


FIGURE 3 Contour plot of the likelihood surface for the *Plethodon* data (Figure 1). The maximum likelihood estimate, the peak of the surface, is marked by an X. The lines surrounding it approximate to the 90, 95, and 99% confidence limits, respectively. (After Nee, S., Holmes, E. C., May, R. M., and Harvey, P. H. (1995). Estimating extinction from molecular phylogenies. In *Estimating Extinction Rates* (J. H. Lawton and R. M. May, eds.) Oxford University Press, Oxford, United Kingdom, pp. 164–182.)

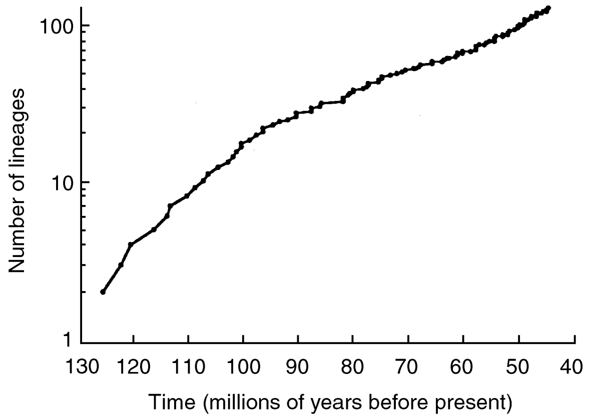


FIGURE 4 A lineages-through-time plot for the molecular phylogeny of birds. All families are represented, that is, all lineages that were present about 45 million years ago and have left descendants in the present day. Accordingly, we should not expect the leveling off of the curve under a constant-rate process: either speciation rate has been decreasing over time or extinction rate has been increasing. There is no evidence for sharp kinks in the curve, which would provide evidence for mass extinction events. For example, the mass extinction at the end of the Cretaceous about 65 million years ago is thought to have resulted in the loss of about 75% of species in many taxa. (After Nee, S., Mooers, A. Ø., and Harvey, P. H. (1992). The tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **89**, 8322–8326.)

reveals two aberrant clades that were speciating at a higher rate when that for others was slowing down. The two radiations resulted in the Passeriformes (seed-eating birds) and the Cicconiformes (mainly shorebirds). The fossil record of both groups is poor, but it is clear that the radiations occurred at a time when landmasses were warmer and shorelines longer than previously. Though cause and effect may be inferred, the unique events described leave the way open for other possible explanations. One way to strengthen interpretations is to show similar increases in biodiversity associated with the similar environmental changes on repeated occasions. If it could be shown, for example, that bursts in speciation among shorebirds were repeatedly associated with increases in the length of shoreline, it would be less likely that a third unknown environmental factor was responsible for changes in shorebird speciation rates.

III. KEY INNOVATIONS, SISTER TAXON ANALYSIS, AND THE ORIGINS OF ADAPTIVE RADIATIONS

It is not necessary that increases in biodiversity occur in response to environmental changes: so-called key innovations may evolve that allow the invasion of currently unoccupied niches. Liem (1973, 1980) suggested that one such innovation is responsible for the much higher rates of speciation among cichlid fishes than among their sister taxa. Liem pointed out that a small shift in position of a single muscle attachment ultimately allowed the pharyngeal bones of cichlids to manipulate their prey items while still holding them. As a consequence, the premaxillary and mandibular jaws were freed to evolve along new routes that did not involve manipulating prey. This, the argument runs, allowed cichlids to evolve a whole new diversity of feeding mechanisms and there seems little doubt that the adaptive radiation of cichlids in African lakes, in the face of competition from other fish families, resulted from their evolved diversity of feeding mechanisms. However, the problem of nonreplication occurs: Lauder (1981) correctly pointed out that any other characteristic common to cichlid fishes but that differs from a sister group could, in principle, be the key innovation that resulted in high speciation rates. Conclusions would be strengthened if the same key innovation had been associated with high species diversity in several taxa.

The first example of a statistically supported successful sister group comparison showing a repeated correlate of high species diversity comes from the Mitter *et al.* (1988) analysis of phytophagy, the habit of feeding on vascular plants, among insects. There are considerable barriers to the evolution of phytophagy. Behavioral and morphological adaptations are required to reduce the risk of desiccation, to remain attached to host, and to deal with low-nutrient food. It had been argued that, once those barriers were overcome, the diversity of phytophages would be promoted by both the great diversity of plant species and of plant parts that could not be exploited by a single species, and by the absence of competitors. Mitter *et al.* performed 13 sister taxon comparisons between the numbers of species in a phytophagous clade versus its nonphytophagous sister clade. Since sister clades originated at the same time, they have had identical times for diversification. In 11 of the 13 comparisons, there were more species in the phytophagous clade than in the sister group, and in each of those cases the difference was greater than twofold.

Sister taxon comparisons using newly established molecular phylogenies are beginning to allow the first real statistical tests of many long-standing hypotheses. For example, Darwin (1871) suggested that sexual selection by female choice might increase the rate of reproductive divergence between populations, and thereby increase the speciation rate of a clade. More than a century later, Barraclough *et al.* (1995) tested the idea using Sibley and Ahlquist's phylogeny of passerine birds. Since it is generally accepted that mate choice is responsible for the evolution of sexual dichromatism in birds, Barraclough *et al.* were able to use sexual dichromatism as a measure of the importance of mate choice for a species. In significantly more than 50% of sister taxon comparisons, the clade with the higher proportion of sexually dichromatic species was more speciose than its sister clade, thereby supporting Darwin's suggestion.

IV. EVOLUTION WITHIN LINEAGES: THE COMPARATIVE METHOD

It is clear that some characteristics of species can promote speciation itself. But species differ from each other in many ways: morphologically, anatomically, physiologically, behaviorally, and by life history. Making sense of the reasons for that diversity is the job of comparative biologists.

A. Fewer Causes Than Characters

Traits have not evolved independently from each other, and particular characteristics of the environment select for whole suites of traits. For example, among many mammalian orders such as the primates, females tend to be spatially distributed in relation to food resources. If food is in large clumps, females live in large groups, but if food is evenly distributed, females are often territorial. Males distribute themselves in response to female grouping patterns, having been selected to monopolize mating access to as many females as possible. When females live alone in territories, males follow suit and share territorial defense (monogamous, as in gibbons). If females live in small groups, then a single male may be able to defend a group of females from mating access by other males (single-male societies). But if females live in groups that are sufficiently large that they cannot be defended by a single male, then more than one male may join the group (multimale societies). Among monogamous species, the males and females have similar weaponry, such as canines, having been selected equally to defend their territory. In single-male and multimale species, males have enlarged canines relative to females of the species, so that they can defend the group of females, or at least any that are in estrus, against mating access from other males. However, in multimale species, there is always the potential for females to be mated by more than one male; males in such species have markedly enlarged testes as a consequence of selection to produce more sperm per ejaculate, thereby increasing the chances of success in sperm competition. Males of monogamous and single-male species are not subject to selection for sperm competition. Here, then, the distribution of resources influences differences in social behavior, weaponry, and testes size. Indeed, if we know a male primate's testes and relative canine sizes, it is possible to say whether he belongs to a monogamous species (both small), a single-male species (large canines, small testes), or a multimale species (both large).

B. Identifying Independent Evolutionary Events

As mentioned in the previous section, there are many fewer explanatory factors than there are variable traits across species. How do we identify the factors that are responsible for the patterns of biodiversity that we see today? Species may share traits in common for two evolutionary reasons: they inherited them from a common ancestor or their ancestors independently evolved the traits. Closely related species share more traits through common ancestry, whereas more distantly related species share more through convergent evolution. This does not necessarily mean that traits cannot evolve rapidly in response to selection. For example, when a new niche appears, the species to successfully invade that niche is most likely to be one that came from a similar niche elsewhere. Very minor changes in phenotype may be sufficient to adapt the new species to its new home. However, the fact that closely related species are subject to similar selective pressures does mean that cross-species correlations can be misleading. For example, in a sample of vertebrates, we might find that those with feathers laid eggs while those with fur gave birth to live young. If the sample contains just birds and mammals, we should be wary of saying that the two sets of characters had responded to the same selective pressure. The way forward is to identify evolutionary independent origins of traits and niche occupancy. For example, we know that on many separate occasions bright coloration among insects has evolved together with distastefulness to predators. Because the two traits seem to have evolved in concert, and there is good reason to expect distasteful prey to advertise the fact to predators, we are more confident that they are part of the same adaptive complex.

For continuously varying characters, such as body size, a particularly elegant technique was developed by Felsenstein (1985) that allows comparative biologists to partition out independent evolutionary change and seek correlated evolution among traits. The key realizations are (1) that differences between each pair of sister taxa in a phylogeny evolved independently, and (2) likely ancestral character states could be estimated if characters had evolved according to a Brownian motion model of evolution. To determine whether characters had evolved independently of each other, it is simply necessary to plot sister taxa differences against each other. If one character evolved independently of another, differences between sister taxa would not be correlated (Fig. 5).

An example of the use of independent contrast analyses is Kelly and Woodward's (1995) analysis of the correlates of variation in carbon isotope composition ($\partial^{13}C$) among plant species. Three cross-species correlates of $\partial^{13}C$ had previously been identified: altitude, latitude, and growth form. When the data were analyzed using phylogenetically independent contrasts, the correlation with latitude dropped out, and when altitude was controlled for there was no correlation with growth form. Alternative interpretations of the reasons for species differences in $\partial^{13}C$ were thereby reduced in number. Indeed, it was concluded that differences in atmospheric composition (CO_2 and O_2 partial pressures) may be sufficient to explain observed differences in carbon isotope composition.

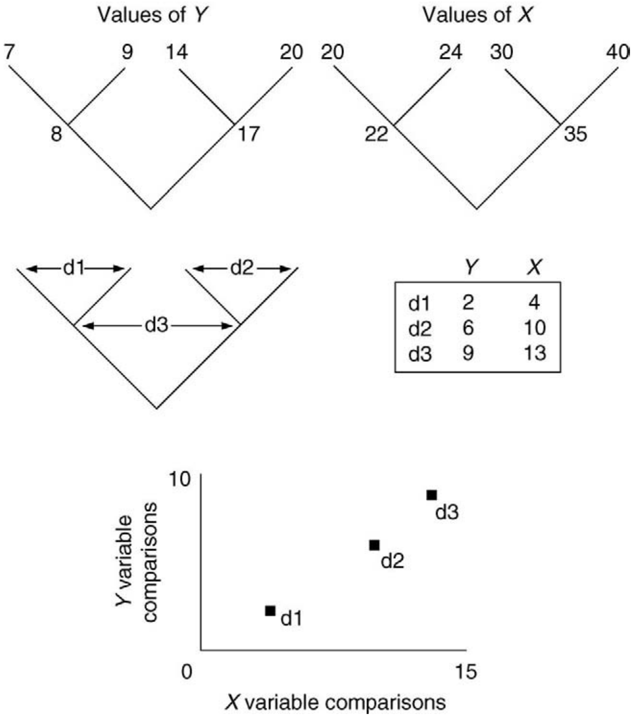


FIGURE 5 The comparative method of independent contrasts. Comparisons d1, d2, and d3 are independent of each other under a Brownian motion model of character evolution. (After Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, United Kingdom.)

See Also the Following Articles

ADAPTIVE RADIATION • BIODIVERSITY, ORIGIN OF • CLADOGENESIS • EVOLUTION, THEORY OF • EXTINCTION, RATES OF • FOSSIL RECORD • PHYLOGENY • SPECIATION, PROCESS OF

Bibliography

Barraclough, T. G., Harvey, P. H., and Nee, S. (1995). Sexual selection and taxonomic diversity in passerine birds. *Proc. Roy. Soc. London, Ser. B* **259**, 211–215.

Barraclough, T. G., Barclay, M. V. L., and Vogler, A. P. (1998). Species richness: Does flower power explain beetle-mania? *Current Biol.* **8**, 843–845.

Benton, M. J. (1993). *The Fossil Record 2.* Harper Collins/Academic, London.

Benton, M. J., and Storrs, G. W. (1996). Diversity in the Past: Comparing cladistic phylogenies and stratigraphy. In *Aspects of the Genesis* *and Maintenance of Biological Diversity* (M. E. Hochberg, J. Clobert, and R. Barbault, eds.). Oxford University Press, Oxford, United Kingdom.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex.* Murray, London.

Farrell, B. D. (1998). "Inordinate fondness" explained: Why are there so many beetles? *Science* **281**, 555–559.

Felsenstein, J. (1985). Phylogenies and the comparative method. *Amer. Naturalist* **125**, 1–15.

Harvey, P. H., May, R. M., and Nee, S. (1994). Phylogenies without fossils. *Evolution* **48**, 523–529.

Harvey, P. H., and Pagel, M. D. (1991). *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, United Kingdom.

Kelly, C. K., and Woodward, F. I. (1995). Ecological correlates of carbon isotope composition of leaves: A comparative analysis testing for the effects of temperature, CO_2, and O_2 partial pressures and taxonomic relatedness on $\partial^{13}C$. *J. Ecol.* **83**, 509–515.

Lauder, G. V. (1981). Form and function: Structural analysis in evolutionary morphology. *Paleobiology* **7**, 430–442.

Liem, K. F. (1973). Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws. *Systematic Zool.* **22**, 425–441.

Liem, K. F. (1980). Adaptive significance of intra- and interspecific differences in the feeding repertoires of cichlid fishes. *Amer. Zool.* **20**, 295–314.

May, R. M. (1996). Introduction. In *Aspects of the Genesis and Maintenance of Biological Diversity* (M. E. Hochberg, J. Clobert, and R. Barbault, eds.). Oxford University Press, Oxford, United Kingdom.

Mitter, C., Farrell, B., and Wiegmann, B. (1988). The phylogenetic study of adaptive zones: Has phytophagy promoted diversification? *Amer. Naturalist* **132**, 107–128.

Nee, S., Holmes, E. C., May, R. M., and Harvey, P. H. (1995). Estimating extinction from molecular phylogenies. In *Estimating Extinction Rates* (J. H. Lawton and R. M. May, eds.) Oxford University Press, Oxford, United Kingdom, pp. 164–182.

Nee, S., Mooers, A. Ø., and Harvey, P. H. (1992). The tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **89**, 8322–8326.

Sibley, C. G., and Ahlquist, J. E. (1990). *Phylogeny and Classification of Birds.* Yale University Press, New Haven, Connecticut.

BIODIVERSITY, ORIGIN OF

M. E. J. Newman and G. J. Eble
Santa Fe Institute

GLOSSARY

Cambrian explosion A period of about 25 million years during the Cambrian, when the diversity of multicellular life greatly increased in a relatively short time.

K-T event A notable large-scale extinction event occurring about 65 million years ago, at the boundary of the Cretaceous and the Tertiary, involving the extinction of about 70% of extant species, including the dinosaurs and many other land-dwelling vertebrates.

monograph effect A term for the fact that the apparent existence of intensive evolutionary activity in a given interval actually reflects the fact that researchers have thoroughly studied a particular section of the fossil record which reveals this activity, while other such investigations elsewhere might just as well have produced the same number of previously unknown species.

Pull of the Recent A term reflecting the fact that recently formed rock is more accessible and offers better preserved fossils; this may produce a bias toward the conclusion that diversity has increased toward the present time.

taxon sorting The argument that in certain taxa, species undergo relatively rapid speciation and extinction, while in others the species speciate and die out relatively slowly; over time this difference will mean that the highly volatile families become extinct more quickly than the less volatile ones.

A GOOD FOSSIL RECORD OF LIFE ON EARTH is available since the beginning of the Cambrian Period about 545 million years ago. Over the interval between then and now, there has been a substantial increase in biodiversity interrupted by many extinction events which caused sharp but relatively brief dips in biodiversity. Modern-day biodiversity is close to the highest it has ever been.

Life began on Earth about 3.5 billion years ago in the Archean Eon, probably with the appearance first of self-reproducing RNAs and later of prokaryotic single-celled organisms. No actual fossils date from this far back, and so our knowledge of Archean life is sketchy at best. Approximately 2.5 billion years ago the last stocks of elemental iron in the earth's crust were converted to oxides as a result of photosynthesis by cyanobacteria, and the planet's atmosphere changed from reducing to oxydizing, making oxygen-breathing life possible. The Proterozoic Eon which followed saw the first appearance of eukaryotes (organisms with nucleated cells and mitochondria), sexual reproduction, and multicellular organisms, although the precise timing and even the order of these innovations are disputed. The earliest firm evidence of metazoan multicellularity dates back to about 575 million years ago.

The end of the Proterozoic at the Vendian–Cambrian boundary about 545 million years ago marks the start of the Cambrian explosion, a period of approximately 25 million years during which, for unknown reasons, multicellular life underwent a period of extraordinary diversification, producing a multitude of new evolu-

Encyclopedia of Biodiversity, Volume 1

tionary lineages in a comparatively short time, including almost all of the major metazoan body plans seen today as well as many which have become extinct. The fossil record from the Cambrian onward—an interval known as the Phanerozoic Eon—is good by comparison with that of the Precambrian largely because multicellular organisms, whose fossils are easier to find and study than those of single-celled organisms, became numerous.

The Phanerozoic fossil record consists of about a quarter of a million known species. The vast majority of these are marine species, mostly invertebrates such as mollusks, brachiopods, corals, or foraminifera as well as sea-dwelling plants. Terrestrial (i.e., land-dwelling) organisms are far less well preserved because the depositional regimes under which fossils are formed are less reliable and uniform on land. Even for marine species, there are many biases in the fossil record, among which the following are some of the most important:

1. Deposition can vary greatly from one time to another. Even in the oceans, in which deposition is relatively reliable, there are periods during which preservation is excellent and others in which it is poor. A sudden period of poor preservation can even give the appearance of a large extinction event because the number of preserved organisms decreases substantially.
2. The "Pull of the Recent" is the name given to the greater accessibility of recently formed rock and the better preservation of fossils in these rocks. The Pull of the Recent implies that we have a more complete record of recent times than we do of times long past, and this could produce the misleading impression of an increase in diversity toward the present where none exists. Despite this bias, it is still believed (as discussed later) that there has been a real increase in biodiversity over the course of the Phanerozoic.
3. The "monograph effect" refers to the sudden apparent burst of new species resulting from the attention paid by a particularly zealous researcher or group of researchers to a particular section of the fossil record: The thorough investigation of a short interval can turn up many previously unknown species, making that interval appear to be one of especial evolutionary activity when in fact a similarly thorough investigation of another interval would have turned up just as many species. A similar effect is produced by the discovery of a particularly rich and well-preserved fossil bed dating from one particular time.
4. Although it is now possible to date fossils within a few million years with high confidence, more accurate dating in the fossil record is still problematic. Absolute dates are derived from slowly decaying radioactive isotopes, whereas relative dates of different fossils are derived from stratigraphic evidence, both geological and from fiducial lineages (so-called "index fossils"). Some geographic areas and intervals of time can be dated with high accuracy—a million years or better—but others are not nearly as good.

Despite these biases, the overall pattern of life during the Phanerozoic survives the vagaries of the record and is now reasonably well understood.

Figure 1 is a graph of (marine) biodiversity as a function of time during the Phanerozoic. The horizontal axis measures time before the present, whereas the vertical one measures the number of known families of marine animals at a succession of times. Each point corresponds to one stratigraphic stage; stages are irregular intervals of time of average duration of about 7 million years which are based on widely accepted stratigraphic features. The data are taken from the compilation of marine families by Sepkoski (1992).

One of the most notable features of Fig. 1 is a substantial increase in the number of known families over the course of the Phanerozoic, particularly in the 250 million years since the Late Permian mass extinction. Overall, the number of known families increases by more than a factor of 10 from the earliest Cambrian to the present. In considering this increase, one must first ask whether it is genuine or an artifact of biases in the fossil record. As mentioned previously, there is certainly a bias occurring because of the better preservation and

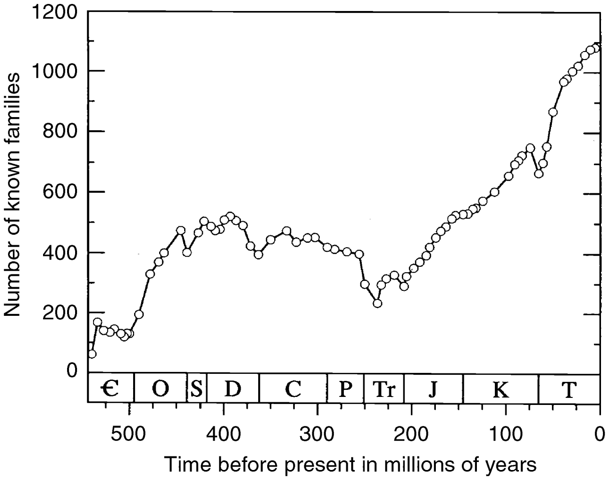


FIGURE 1 The number of known marine families alive over the time interval from the Cambrian to the present (data from a compilation by Sepkoski, 1992).

greater availability of fossils in more recent rocks. There is no doubt that some of the increase apparent in Fig. 1 is a result of this bias. On the other hand, the use of numbers of families as a measure of diversity, rather than species, tends at least partly to remove this bias since we need find only one species belonging to a certain family in order to establish the existence of that family at any given time. Therefore, even during periods of relatively poor species sampling, a family with a sufficiently large number of species can still contribute to our estimate of diversity. Furthermore, there are plausible biological reasons to believe in a real diversity increase: The increase seems to some extent to be the result of the evolution of organisms into new ways of life. The spread of life onto the land in the Silurian and Devonian is a particularly prominent example of this, but among marine organisms the trend is also clear. It seems reasonable to suppose that as life spreads into new environments, the total number of species (or families) which the planet can support will also increase.

Many models have been put forward for diversity increases in the fossil record. Benton (1995) suggested that the increase is an exponential one, indicating perhaps a stochastically constant rate of creation of new taxa from existing ones, with no detectable limit to ultimate levels of diversity. A more detailed analysis has been given by Sepkoski (1981), who suggested that the diversification of life occurred in three distinct phases, each typified by vigorous turnover within a particular subset of species or "fauna." The first phase, during the Cambrian, was dominated by groups in which the rate of origination of new species was relatively high, such as trilobites, inarticulate brachiopods, and monoplacophorans. The second phase, lasting from the Ordovician to the end of the Permian, was dominated by groups with intermediate origination rates, such as crinoids, articulate brachiopods, and cephalopods. Finally, there was a third phase, lasting from the beginning of the Triassic up until the Recent, which was dominated by bivalves, gastropods, echinoids, and marine vertebrates—groups which have relatively low rates of origination. Sepkoski (1984) suggested that within each of the three faunas diversification proceeds initially exponentially but eventually saturates, implying the existence of equilibrium "carrying capacities" for the global ecosystem. The superimposition of this "logistic" growth within each fauna gives rise to the observed pattern of near-monotonic diversity increase in Fig. 1. A logistic diversification model was also explored by Courtillot and Gaudemer (1996) but without allusion to distinct faunas.

An important feature of the pattern of biodiversity shown in Fig. 1 which can be explained in terms of the logistic growth model is the apparent plateau in diversity throughout the greater part of the Paleozoic, from the Late Ordovician to the end of the Permian. [The same pattern is also present in genus-level data (Sepkoski, 1997), although there is in general more fluctuation in the numbers of genera so the plateau is not as clear.] One explanation for this plateau is that it represents the saturation phase of the logistic diversification of Sepkoski's second fauna, which continues mostly uninterrupted for approximately 200 million year until the Late Permian mass extinction. This is not the only possibility, however; it has also been suggested that the plateau is more the result of high extinction rates in the later Paleozoic than it is the result of low origination rates (Stanley, 1999). In both cases, however, the ultimate limit on diversity is presumed to be some effective carrying capacity of the ecosystem.

Another notable feature of Fig. 1 is a dip in diversity at several points, particularly at the ends of the Devonian (D), Permian (P), and Cretaceous (K) periods. These dips correspond to major mass extinctions that can be seen more clearly in Fig. 2, which shows the percentage of extant families of marine animals becoming extinct in each stratigraphic stage of the Phanerozoic. The data are again taken from the compilation by Sepkoski (1992). We normally distinguish five major mass extinctions during the Phanerozoic. These "big five" are marked with arrows in Fig 2. A sixth period of heightened extinction is visible in the Cambrian but is thought to be primarily an artifact of sampling biases

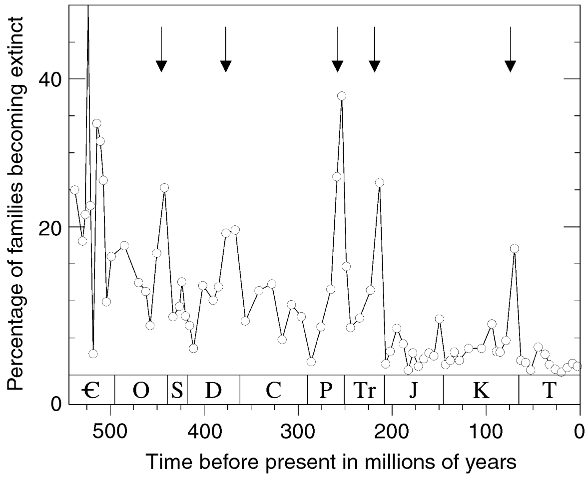


FIGURE 2 Estimated extinction of marine animals in families per stratigraphic stage since the Cambrian as a percentage of the total number of families in existence. The arrows indicate the positions of the "big five" mass extinction events discussed in the text.

rather than a real extinction event. Many smaller extinction peaks are also apparent in Fig. 2 but are of secondary importance. The basic features of the big five events are as follows.

The end-Ordovician event about 440 million years ago appears to have occurred in two bursts, separated by about 1 million years, which between them wiped out about 85% of then living species. The event was confined to marine species since multicellular life had not yet colonized the land. Particularly affected were brachiopods, bivalves, echinoderms, bryozoans, and corals. The immediate cause of extinction appears to have been the continental drift of a significant land mass into the south polar region, causing a global temperature decrease, glaciation, and consequent lowering of the sea level, which destroyed species habitats around the continental shelves. The sea level rose again with the end of the glacial interval about 1 million years later and caused a second burst of extinction.

The Late Devonian extinction approximately 360 million years ago is complex and poorly understood. It is probably in fact composed of many separate events (as many as seven) spread over about 25 million years, including particularly notable extinctions at the ends of the Givetian, Frasnian, and Famennian stages. Overall, about 80% of living species died out in the Late Devonian. Particularly hard hit were corals, brachiopods, bryozoans, ammonoids, and fish. The causes of these extinctions are unclear. The leading theories suggest that changes in sea level and ocean anoxia, possibly triggered by global cooling or oceanic volcanism, were most likely responsible, although the impact of an extraterrestrial body such as a comet has also been considered.

The Late Permian extinction approximately 250 million years ago was the largest extinction event of all time, killing approximately 95% of marine species and about 70% of land-dwelling ones. Like the end-Ordovician event, it seems to have been composed of two bursts, separated in this case by an interval of about 10 million years, with the second being the larger of the two. Notable extinction happened again among brachiopods, ammonoids, and corals as well as gastropods, echinoderms, and, unusually, insects. Despite an enormous amount of research of the subject, the causes of the Late Permian event are still a subject of debate. It is clear, however, that the sea level rose during this period, levels of oxygen in the oceans were low, and carbon dioxide levels were high. There is some suggestion that a cometary impact may have been involved, or a shift in ocean circulation driven by climate change, or carbon dioxide and sulfur release following large-scale volcanic activity. The Late Permian event had a profound effect on the terrestrial ecosystem which is still being felt today, a quarter of a billion years later. A particularly notable example among marine faunas is that of the bivalves, a relatively minor group during the Paleozoic that took advantage of the ecological vacuum left by the extinction to establish a solid grip on shallow-water environments, leading to their dominance over the previously very successful brachiopods.

The end-Triassic extinction approximately 210 million years ago is probably the most poorly understood of the big five extinction events. It appears to have killed about 80% of species then living, either in one burst or possibly in two separated by about 20 million years. Major extinction is observed particularly among ammonoids, bivalves, gastropods, and brachiopods. Leading theories of the causes of the end-Triassic event are ocean anoxia, massive volcanism, or possibly a bolide impact.

The end-Cretaceous event, usually called the Cretaceous–Tertiary or K–T event, has attracted the most popular interest of any extinction because it saw the end of the dinosaurs, but it was in fact the smallest by quite a wide margin of all the big five. The K–T event appears to have been a single pulse of extinction approximately 65 million years ago which wiped out about 70% of all species then living. In addition to the dinosaurs, it extinguished many other land-dwelling vertebrates, especially large-bodied ones, along with large numbers of (marine) bivalves, gastropods, and foraminifera. The proximal cause of the K–T event was almost certainly the impact of a large comet or meteor near the present site of the town of Puerto Chicxulub on the Yucatán peninsula in eastern Mexico, with an associated drop in sea level and possibly short-term cooling or heating or acid rain. It has also been noted that average extinction rates were higher than the historical average for some time in advance of the K–T boundary, indicating that long-term environmental change may also have played a part.

Table I summarizes the fraction of families and species killed in each of the big five mass extinctions. In fact, the fraction of species killed cannot be measured directly from the fossil record since the preservation of individual species is too unreliable; the majority of species appear in only one stratigraphic stage so that no statistically significant estimate of their origination or extinction time can be made. The fraction of families killed can be calculated with reasonable certainty: Many families have reasonable coverage in the fossil record so that good estimates can be made of origination and extinction times. Estimates of species kill are then ex-

TABLE I

Extinction Intensities at the Family and Species Level for the Big Five Mass Extinctions of the Phanerozoic[a]

Extinction	Family loss (observed) (%)	Species loss (estimated) (%)
End Ordovician	26	84
Late Devonian	22	79
Late Permian	51	95
End Triassic	22	79
End Cretaceous	16	70

[a] Estimates of family extinction are obtained from directed analysis of the fossil record, whereas species loss is inferred using a statistical technique called "reverse rarefaction." Data are taken from Jablonski (1995).

trapolated from the family-level estimates by a technique known as reverse rarefaction which works as follows. First, one denotes the probability of extinction of a particular species within a family as p. If the family contains only one species, then p is also the probability of extinction of the family. If it contains two species, however, then the probability of extinction of the family is p^2. In general, if the family contains n species, then the probability of its extinction is p^n. Then one averages this over many families to find the average probability of extinction of a family. In order to perform the average, we need to know the distribution of the sizes of families. Following Raup (1979), data from living species are usually used for this purpose. Calculating the average probability of family kill gives us a "rarefaction curve" (Fig. 3) for the fraction of families or genera killed as a function of the fraction of species killed. "Reverse rarefaction" is the name given to the use of this curve "backwards" to estimate what fraction of species must have been killed in order to produce an observed level of family kill. This is the method which is used to calculate the data in Table I.

Another overall feature of the extinction record shown in Fig. 2 is a decline in the level of background extinction (i.e., not mass extinction) during the Phanerozoic. Like the increase in standing diversity discussed earlier, it is important to establish whether this decline is a real effect or the result of some bias in the record. It is difficult to imagine that a decrease in the number of families becoming extinct could arise from better preservation toward the Recent. Rather, one would expect the reverse since, if extinction were actually approximately constant but more fossil families were preserved, one would expect an increase in apparent extinction. What is the explanation for the decrease? Many hypotheses have been put forward.

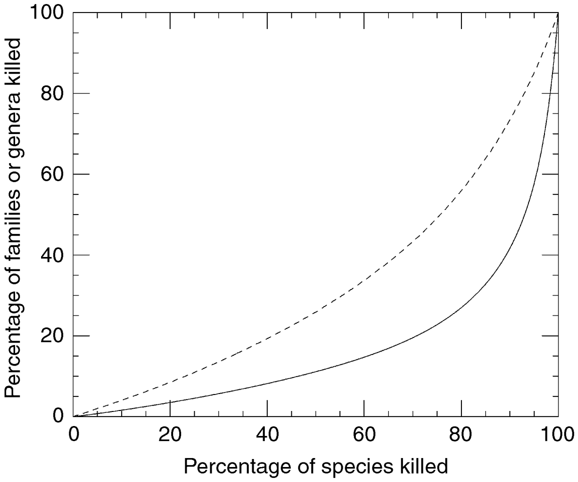


FIGURE 3 Rarefaction curves for the extinction of families (solid line) and genera (dashed line) calculated for echinoderms by Raup (1979). The curves indicate what percentage of families or genera are expected to become extinct during an event which kills a given percentage of species. Used in reverse, the curves also allow us to estimate what percentage of species became extinct in a particular event, given an observed percentage of family or genus kill.

Flessa and Jablonski (1985) suggested that the decrease could be a result of increases in the average number of species per family toward the Recent. If families increase in size (as seems to have been the case), then the probability of extinction, which was denoted as p^n, will decrease as n, the number of component species, becomes larger. A related suggestion has been made by Boyajian (1986), who proposed that families "age" over time, meaning that they accumulate more species (on average) the longer they live and therefore become resistant to extinction. If the number of such aged families increased over time, this could explain the decrease in extinction rates. This explanation receives some support from the fossil record, which indicates that families have indeed tended to live longer in recent times than they did in the early Phanerozoic. The average lifetime of a family in the Paleozoic is estimated to be about 30 million years but increases to about 80 million years in the post-Paleozoic.

Gilinsky and Bambach (1987) and Gilinsky (1994) made a slightly different argument that the decrease is a "taxon sorting" effect. Suppose that taxa (e.g., families) vary in volatility. That is, the species within some of them undergo comparatively rapid speciation and extinction and turnover is high, whereas the species within others speciate and die out relatively slowly. For the highly volatile families, the time to extinction of the entire family, as a result of extinction of all its members, will be shorter than that for less volatile ones by precisely the ratio of the respective volatilities, meaning that the highly volatile families become extinct more quickly. Over the course of a long period of time, therefore, the families that remain will tend to be the less volatile, longer lived ones. Taxon sorting in effect produces a selection pressure at the family or other higher taxonomic level. Another possible explanation for the increase in the lifetimes of taxa and decrease in extinction rates is that species have really become more highly evolved toward the Recent and are better adapted to survive (Raup and Sepkoski, 1982). This explanation, however, is not a widely accepted one.

Although extinction has received the lion's share of paleontological attention in recent years, origination is equally important in the study of Phanerozoic biodiversity. We have already discussed origination in the context of the "three faunas" view of the diversification of life. Figure 4 shows a detailed plot of the measured pattern of origination of families during the Phanerozoic. Many prominent features are visible. The major peaks in the Cambrian period are a reflection of the burst in metazoan diversity of the Cambrian explosion. Since the plot is one of origination rates as a fraction of standing diversity, however, the size of these peaks is partly just a result of low Cambrian diversity (Fig. 1). Another burst of origination is visible in the Ordovician. This burst was responsible for the corresponding major increase in diversity shown in Fig. 1. After the Ordovician, origination is more subdued, although substantial surges did take place in the Triassic, Jurassic, Cretaceous, and during the Cenozoic.

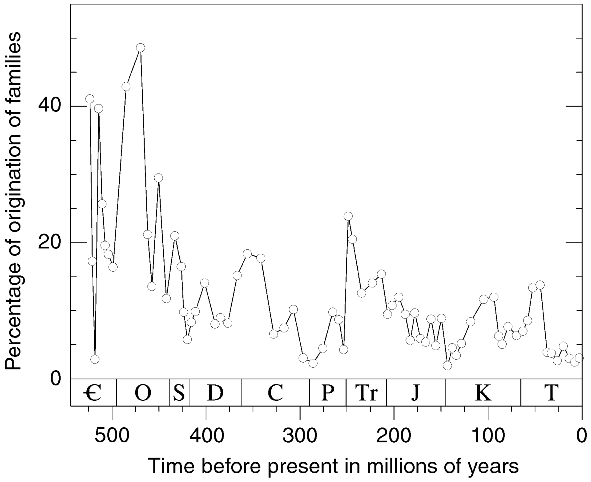


FIGURE 4 Estimated origination of marine animals in families per stratigraphic stage since the Cambrian as a percentage of the total number of families in existence.

On the whole, origination tends to vary less in magnitude and to be less episodic than extinction (Foote, 1994). Nonetheless, it appears to be bursts of origination, rather than declines in extinction rates, which account for most of the major evolutionary radiations. The basic features of the Ordovician, Jurassic–Cretaceous, and Cenozoic radiations are as follows.

Origination during the Ordovician occurred particularly among groups such as brachiopods, bryozoans, and mollusks, increasing the dominance of these groups while groups such as trilobites shrank in importance. Although phyla had basically ceased to be produced by the Ordovician, the period led to substantial production of body plans at the level of classes and orders.

The Jurassic–Cretaceous radiation produced substantial increases in the diversity of gastropods, bivalves, echinoids, and fishes—groups that were relatively minor during the Paleozoic. Ecological expansion in the form of the appearance of fundamentally new modes of life (Bambach, 1985) and the increasing occurrence of predation (Vermeij, 1977) have been advanced as important contributors to this pattern.

The Cenozoic radiation is certainly in part a product of the Pull of the Recent, but there are also genuine bursts of origination in the Paleocene and Eocene. In the marine realm, the diversity of veneroid bivalves, neogastropods, irregular echinoids, gymnolaemate bryozoans, and fishes increased substantially, whereas in the terrestrial realm the mammal and angiosperm radiations are particularly prominent.

Although mass extinctions are usually explained as the results of various kinds of exogenous stress on the ecosystem, such as climate change or bolide impact, originations seem to be more often the result of biological phenomena such as evolutionary innovations or changes in the structure of ecological assemblages. However, as a close inspection of Figs. 1, 2, and 4 reveals, origination is often intensified in the aftermath of mass extinction events, suggesting that physical disturbances do have an indirect effect on origination, probably through removal of incumbent species, through selectivity in extinction (Jablonski, 1989), and through particular patterns of rediversification of groups and repopulation of ecospaces during the recovery period following extinction. Recoveries typically take a few million years to run their course and reestablish previous levels of diversity and ecological heterogeneity (Erwin, 1998). Some recoveries do not begin immediately after an extinction event; there is a lag, perhaps as long as a few million years, probably the result of lingering environmental stresses and the proliferation of opportunistic species. A paradigmatic example of recovery is the Triassic recovery which followed the end Permian (Fig. 4). A conspicuous spike in origination is observed beginning about 5 million years after the end of the extinction and continuing for about another 5 million years.

A long-term decline in origination over the Phanerozoic is as clear a feature of the fossil record as is the decline in extinction. Taxon sorting of the kind described earlier in the context of extinction is a plausible explanation (Gilinsky, 1994). Alternatively, the decline in origination may reflect a genuine decrease in the rate of evolutionary innovation as a result of either ecological or developmental restriction (Gilinsky and Bambach, 1987; Eble, 1999).

It is also instructive to examine origination rates at the level of higher taxa such as classes and orders. Origination at these levels reflects real innovations (to the extent that they are visible in the fossil record) rather than just sheer numbers of species. Indeed, there is no reason why the two need necessarily be correlated. During the Cambrian, for example, changes in total diversity (e.g., measured by the number of families) are modest, but it is in the Cambrian that one finds the most significant generation of new body plans—the "Cambrian explosion." This aspect of biodiversity having to do not with numbers of evolving entities but with their distinction is often referred to as "morphological diversity," "biodisparity," or simply "disparity" (Foote, 1997). Significant discrepancies between biodiversity and biodisparity can exist, such as faster morphological diversification early on in the evolution of clades, despite low levels of taxonomic diversity, and later deceleration despite continued increases in diversity. Further understanding of biodiversity in the fossil record may hinge on a better understanding of such patterns.

See Also the Following Articles

DIVERSITY, ORGANISM LEVEL • EUKARYOTES, ORGIN OF • FOSSIL RECORD • MASS EXTINCTIONS, CONCEPT OF • MASS EXTINCTIONS, NOTABLE EXAMPLES OF • ORIGIN OF LIFE, THEORIES OF • PALEOECOLOGY • SPECIES DIVERSITY, OVERVIEW

Bibliography

Bambach, R. (1985). Classes and adaptive variety: The ecology of diversification in marine faunas through the Phanerozoic. In *Phanerozoic Diversity Patterns: Profiles in Macroevolution* (J. W. Valentine (Ed.). Princeton Univ. Press, Princeton, NJ.

Benton, M. J. (1995). Diversification and extinction in the history of life. *Science* **268**, 52–58.

Boyajian, G. F. (1986). Phanerozoic trends in background extinction: Consequence of an aging fauna. *Geology* **14**, 955–958.

Courtillot, V., and Gaudemer, Y. (1996). Effects of mass extinctions of biodiversity. *Nature* **381**, 146–148.

Eble, G. J. (1999). Originations: Land and sea compared. *Geobios* **32**, 223–234.

Erwin, D. H. (1998). The end and the beginning: Recoveries from mass extinctions. *Trends Ecol. Evol.* **13**, 344–349.

Flessa, K. W., and Jablonski, D. (1985). Declining Phanerozoic background extinction rates: Effect of taxonomic structure? *Nature* **313**, 216–218.

Foote, M. (1994). Temporal variation in extinction risk and temporal scaling of extinction metrics. *Paleobiology* **20**, 424–444.

Foote, M. (1997). The evolution of morphological diversity. *Annu. Rev. Ecol. Syst.* **28**, 129–152.

Gilinsky, N. L. (1994). Volatility and the Phanerozoic decline of background extinction intensity. *Paleobiolog* **20**, 445–458.

Gilinsky, N. L., and Bambach, R. K. (1987). Asymmetrical patterns of origination and extinction in higher taxa. *Paleobiology* **13**, 427–445.

Jablonski, D. (1989). The biology of mass extinction: A palaeontological view. *Philos. Trans. R. Soc. B* **325**, 357–368.

Jablonski, D. (1995). Extinctions in the fossil record. In *Extinction Rates* (J. H. Lawton and R. M. May, Eds.). Oxford Univ. Press, Oxford.

Raup, D. M. (1979). Size of the Permo-Triassic bottleneck and its evolutionary implications. *Science* **206**, 217–218.

Raup, D. M., and Sepkoski, J. J., Jr. (1982). Mass extinctions in the marine fossil record. *Science* **215**, 1501–1503.

Sepkoski, J. J., Jr. (1981). A factor analytic description of the Phanerozoic marine fossil record. *Paleobiology* **7**, 246–267.

Sepkoski, J. J., Jr. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology* **10**, 246–267.

Sepkoski, J. J., Jr. (1992). A compendium of fossil marine animal families, 2nd edition. *Milwaukee Public Mus. Contrib. Biol. Geol.* **83**.

Sepkoski, J. J., Jr. (1997). Biodiversity: Past, present, and future. *J. Paleontol.* **71**, 533–539.

Stanley, S. M. (1999). The myth of the diversity plateau: High extinction rates, not ecological saturation, stifled the diversification of Paleozoic marine life. *Geol. Soc. Am. Meeting Abstr.* **31**, A337.

Vermeij, G. J. (1977). The Mesozoic marine revolution: Evidence from snails, predators and grazers. *Paleobiology* **3**, 245–258.

BIODIVERSITY-RICH COUNTRIES

José Sarukhán and Rodolfo Dirzo
Instituto de Ecología, Universidad Nacional Autónoma de México

GLOSSARY

beta-diversity Turnover of species within a small spatial scale resulting from highly differing ecological conditions.

biogeographic regions Major areas of the world where large assemblages of species originated and that differ considerably from other regions.

CONABIO The Mexican National Commission on Biodiversity.

Convention on Biodiversity (CBD) An international convention signed by 175 countries, which originated at the Earth Summit of Rio de Janeiro in 1992.

cultural diversity Variation of races, languages, and habits that define a given culture.

database Assemblage of information organized in tables with a logical structure. It is a fundamental tool used to gather, organize, and analyze biodiversity information.

endemism Confinement of a species or other taxonomic group to a given region, for example, an island or a country.

ERIN The Environmental Resources Information Network of Australia.

INBio The Biodiversity Institute of Costa Rica.

megadiverse An area or a country possessing a much larger proportion of species than would be expected by its extent, latitudinal position, and other factors.

parataxonomist Laymen and peasants trained to collect and to carry out preliminary identification of botanical and zoological specimens. They are widely used by INBio in Costa Rica.

phylum A major category, between kingdom and class, of taxonomic classification.

primary production Amount of energy produced by photosynthetic organisms in a community.

SABONET The South African Botanical Diversity Network of Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia, and Zimbabwe.

BIOLOGICAL DIVERSITY IS NOT HOMOGENEOUSLY DISTRIBUTED AROUND THE GLOBE. The latitudinal variation in biological diversity is an obvious example. However, superimposed on this and other natural trends of spatial variation is the heterogeneous distribution of diversity according to the geographical position of political entities, such as countries. A handful of countries in which biodiversity is particularly overrepresented constitute the so-called megadiversity countries. Obviously there is no reason for biodiversity to follow the artificial boundaries of countries. However, the reality is that such geopolitical entities exist, and they have particular ecological, historical, social, and economic structures that must be taken into account in any attempt to view, value, and conserve global biodiversity.

I. INTRODUCTION

The notion of megadiversity countries was first suggested by the well-known conservation biologist Russell Mittermeier, who developed it with an initial emphasis on tropical primates. Later it was extended to all types of ecosystems and several groups of organisms (Mittermeier *et al.*, 1997). The concept of megadiversity countries is close to that of centers of diversity, which refers to the existence of areas with high biodiversity, particularly large numbers of species and a high concentration of endemic organisms. However, there are important distinctions. Centers of diversity are natural spatial units, and they may be recognized at several scales, such as local, regional, and global. Megadiversity countries, on the other hand, are spatial entities restricted within geopolitical limits and the recognition of variation at different scales can only be made within the countries' artificial boundaries or at the global scale.

It is not surprising that there is a lack of precise numerical coincidence between centers of diversity and megadiverse countries. For instance, at the global scale, the IUCN (World Conservation Union) Plant Conservation Office has recently recognized 234 centers of plant diversity, while the number of megadiverse countries, from the botanical point of view, is only about 10. However, some spatial overlap can be expected: of the total number of centers of plant diversity, close to 50% are located within these 10 megadiversity countries. In addition, the notion of megadiversity country implies several biopolitical connotations. One of the most evident of these is that such countries have a special responsibility to protect such high concentrations of biological resources. That such concentrations of the total biodiversity of the planet are located in a handful of countries constitutes a matter of global as well as national concern.

This article describes the salient facets of biodiversity that determine the existence of megadiverse countries, and then addresses the current threats that this diversity is experiencing. A discussion of relevant sociodemographic traits of megadiverse countries is presented, together with the repercussions for the conservation and management of natural resources. Special attention is given to the large indigenous human populations that inhabit most of these countries. Finally, we explore the factors that limit the capacity of many megadiverse countries to fully know their biological diversity and to be able to utilize that knowledge for the purpose of conserving and managing their natural ecosystems. We examine examples of how some countries have developed institutional efforts to compile, systematize, and utilize biodiversity information in order to convert it into conservation policy.

II. FACETS OF BIODIVERSITY IN MEGADIVERSE COUNTRIES

The first issue when describing the underlying causes for the concentration of biodiversity in specific countries is the establishment of criteria upon which the definition of megadiversity is based. Although several authors have used different approaches to rank areas according to their biodiversity, consideration of species richness predominates in this article, and following Mittermeier *et al.* (1997), megadiversity is defined on the basis of the following facets of biological diversity.

A. Species Richness

We base our assessment in terms of the species of plants and four groups of animals: mammals, birds, amphibians, and reptiles (which are, in general, the best known taxa). Species richness in these groups of organisms, and plant species in particular, can be taken as correlates of species richness of other, lesser-known organisms. For example, species richness of a critical group, insects (at least some groups), is highly dependent on plant species richness. When available, species richness of other groups of organisms was also used.

B. Concentration of Endemisms

The concentration of endemisms refers to the percentage of endemic species present among a country's total resident species. This criterion is also based on the best-known groups of organisms: plants, mammals, birds, amphibians, and reptiles.

C. Diversity of Habitats

The diversity of habitats refers to the diversity of distinct ecosystems, usually considered to be the diversity of vegetation types. Closely related to this component is the concept of beta diversity, which is the spatial turnover of species between nearby or adjacent habitats. High beta diversity frequently results from the heterogeneity of ecological conditions. Commonly, significant species turnover also occurs at regional scales within a given vegetation type.

D. Presence of Tropical Forest Ecosystems

Given that tropical forests, particularly tropical rain forests, are widely recognized as the most species-rich terrestrial ecosystems, their presence has been used as another important facet of megadiversity. This criterion was used only in qualitative terms, with no attempt at quantifying the expanse of tropical forests per country.

E. Presence of Marine Ecosystems

The presence of marine ecosystems was also included because such systems hold the greatest diversity of taxa of higher rank. In particular, marine ecosystems are rich in phyla of animals: of the 33 known animal phyla, 28 are present in the marine realm and 13 are exclusive to it. In contrast, terrestrial ecosystems include 10 phyla, with only one endemic to it. In addition, marine ecosystems contain a great diversity of fish, which was used as a complementary criterion.

F. Cultural Diversity

The cultural diversity criterion addresses the diversity of ethnic groups (or autochthonous languages) that are present in a given country. The relevance of this component stems from the fact that there is a strong association between biological diversity and cultural diversity. Different cultural dimensions in different countries have played a significant role in the modes in which biodiversity is perceived, maintained, used, and appreciated.

III. UNDERLYING CAUSES OF MEGADIVERSITY

The next step is the analysis of the causes, proximal or ultimate, that determine the conjunction of the foregoing attributes in a given country. The predominant underlying causes that determine the existence of the megadiversity countries are latitudinal position, physical factors that increase primary production, historical factors, ecological diversity, and the presence of tropical rain forests.

A. Latitudinal Position

Latitudinal position is correlated with a number of ecological factors; by itself it is not a determining factor of high biodiversity. Rather, it is a proximal correlate of biodiversity: for almost all groups of organisms, species diversity increases toward equatorial latitudes. For example, on land, species richness and latitude are negatively correlated among mammals, birds, and trees (Fig. 1). Likewise, the number of species of ants in local regions increases from about 10 at 60°N latitude to about 2000 in equatorial regions.

It follows, therefore, that countries located at lower latitudes tend to be rich in species of many groups. Table I shows the contrast between tropical and temperate countries in terms of the number of plant and mammal species. Regardless of size, the sample of countries located close to tropical latitudes have plant and mammal diversities that are, on average, two times larger than those of the temperate countries. Evidently the tropical countries of this sample constitute a list of potential megadiversity candidates. Nevertheless, Table I shows that some of the temperate countries (e.g., Australia) reach values of species richness that are comparable to those of tropical countries. This implies that other factors besides latitude may be important correlates or causal factors of species richness. For example, in the case of Australia, its large territorial size and the fact that its northern region has tropical ecosystems account for its richness.

It may also be noted that historical factors could have played an important role in establishing the latitudinal

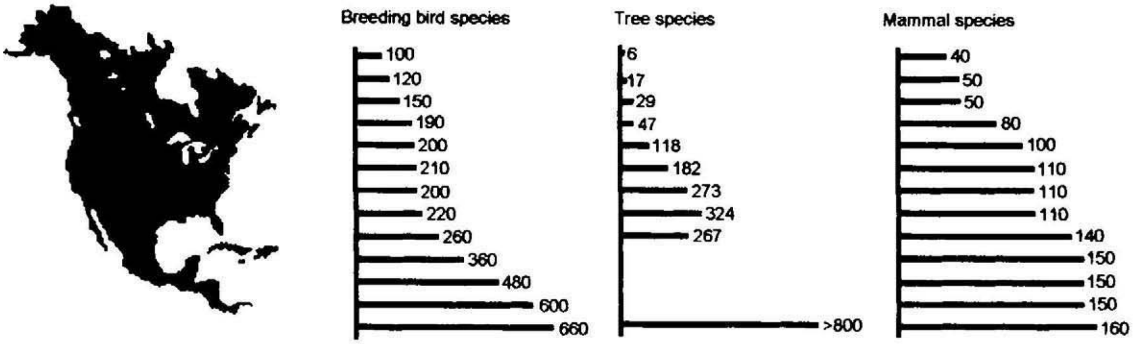


FIGURE 1 Latitudinal gradients of species richness for birds, trees, and mammals in North and Central America. Bars and numbers correspond to species richness according to latitudinal position on the map to the left. (From Briggs, 1995; Fig. 4.5 of Meffe and Carroll, 1997.)

patterns of distribution of biological richness. High latitudes were covered by ice until relatively recently, whereas most equatorial latitudes have been relatively stable from the climatic point of view. This allows for a longer period of speciation and accumulation of taxa at lower than at higher latitudes.

The patterns of decrease in species richness with latitude in terrestrial ecosystems is paralleled in marine systems, although available information relates to other groups of organisms. For example, Arctic waters contain about 100 species of tunicates (sea squirts), temperate seas contain some 400 species, and waters at tropical latitudes include over 600 species. Likewise, the richness of species, genera, and families of bivalve mollusks peaks in tropical regions and declines with increasing latitude (Fig. 2). Although data are not readily available for individual countries, the similarity of latitudinal patterns in terrestrial and marine systems suggests that a similar trend exists on a per country basis, which again indicates that low-latitude countries are likely megadiversity candidates on the basis of marine species richness.

B. Physical Factors of the Environment Leading to Increased Primary Production

In many parts of the world, primary production increases with precipitation, available solar energy, and nutrient concentration (e.g., soil fertility). The highest terrestrial primary production is found in areas with high rainfall and year-round warm temperatures, whereas production decreases with lowered temperatures and the occurrence of frost. In marine settings, primary production increases with nutrient concentra-

TABLE I

Numbers of Plant and Mammal Species in a Selected Sample of Tropical and Temperate Countries

Tropical country	Plants	Mammals	Temperate country	Plants	Mammals
Brazil	50,000–56,000	524	Argentina	9,000	255
Colombia	45,000–51,000	456	Australia	15,638	282
Democratic Republic of Congo	11,000	415	Canada	2,920	163
Ecuador	17,600–21,100	271	Egypt	2,066	105
India	>17,000	350	France	4,500	113
Indonesia	~37,000	515	Japan	4,700	186
Malaysia	15,000	286	Morocco	3,600	108
Mexico	18,000–30,000	450	South Africa	23,420	247
Peru	18,000–20,000	344	UK	1,550	77
Venezuela	15,000–21,070	288	USA	18,956	428

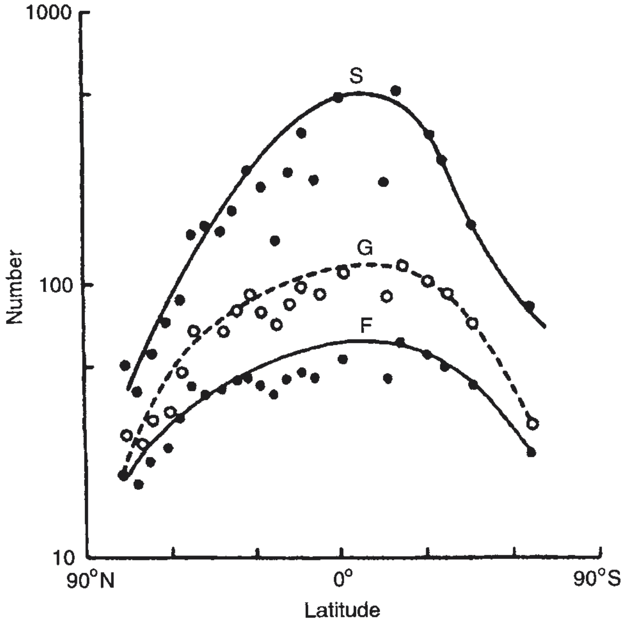


FIGURE 2 Latitudinal variation in richness of species (S), genera (G), and families (F) of bivalve mollusks. (From Stehli *et al.*, 1969; Fig. 4.4 of Meffe and Carroll, 1997.)

tion, as is typical of some coastal zones, and it is lowest in open ocean where nutrient concentration becomes limiting (but apparently this is not so on the bottom of the ocean and at thermal vents). In addition, in many ecosystems, species richness increases with the increment in primary production at low productivity levels, reaches a plateau at moderate levels of production, and then decreases at high levels of primary production. Thus, areas with environmental factors leading to increased primary production have, up to a certain level, a greater resource base that can support a wider range of species and typically have high species richness. As a result, countries with latitudinal and elevation ranges that include tropical and lowland regions, as well as coastal areas where nutrient-rich waters upwell, tend to be megadiverse. Brazil, Peru, Colombia, and Mexico are clear examples of countries where such environmental conditions coincide.

C. Historical Factors

Historical factors are also critical in determining species richness and the concentration of endemisms. Over a long-term geological timescale, the continued rise in species richness is explained by the "provincialization" that accompanied the breakup of Pangaea and, later, Gondwanaland. Following their separation, evolution on each of the newly produced "provinces" (continents) produced many new taxa that were continental endemics and, indeed, much of the increase in species numbers of the Cenozoic era appears to be due to such continental isolation. As a consequence, several countries of South America (Brazil, Colombia), Africa (Democratic Republic of Congo), and Southeast Asia (Indonesia) are extremely rich in species, but they share few species in common because of their separation of many millions of years.

In general, areas that are geologically older have higher species richness than younger areas. Such geologically older areas have had more time to receive species dispersing from other regions of the world and more time for existing taxa to undergo adaptive speciation in response to the local conditions. The effect of geological age on species richness is well exemplified among marine ecosystems. The species richness of reef-forming corals is several times greater in the Indian Ocean and western Pacific than in the Caribbean Sea and adjacent Atlantic Ocean, which are considerably younger than the former. A similar situation is the case of mangrove ecosystems, which typically contain a much higher number of arboreal mangrove species in the Old World than in the Neotropics. Marine species richness of countries associated with these older seas is therefore remarkably high, thus contributing to the existence of megadiverse countries in some of these areas. Australia is a clear example: its Great Barrier Reef contains about 8% of the world's fish species in just 0.1% of the global ocean surface area. The total diversity of fish species in Australia, about 4300, is equivalent to 20% of the world's total.

Regions rich in endemism tend to have experienced major events that caused fragmentation of the ranges of many species at approximately the same place. Such fragmentation events leading to geographic isolation include continental drift, mountain building, and sea level changes. Subsequent to their isolation, taxa may undergo evolutionary radiation in a given locality. For example, isolation and subsequent evolutionary radiation due to continental drift have been fundamental to the generation of a high degree of endemism in the biota of Australia, Papua New Guinea, and Madagascar. Mountain building that led to isolation, explosive radiation, and a high accumulation of endemism is well represented in countries such as Colombia, Peru, and Mexico. Changes in sea levels have been responsible for the remarkable species richness and concentration of endemism in Indonesia and also in Mexico.

Likewise, the remarkable diversity of fishes and other groups of organisms in large tropical lakes and rivers is due to rapid evolutionary radiation in such isolated and usually productive ecosystems. Notable in this respect are countries such as Brazil, Colombia, and Peru.

As expected, islands, owing to their isolation, often have high proportions of endemic taxa. Australia, a continent-island of remarkable size, is one of the world's leaders in the endemism of mammals, birds, reptiles, and amphibians. Nevertheless, because islands frequently have a relatively depauperate biota, such high levels of endemism do not necessarily coincide with high richness of species, except on those islands where other driving factors are of exceptional importance, such as the large archipelago of Indonesia.

Another remarkable aspect related to biogeographic history is the occurrence of distinctive floristic and faunistic assemblages, which define the so-called biogeographic regions or provinces. Six of them have been identified on the basis of the geographic affinity as well as the presence and absence of particular taxa. Such biogeographic provinces reflect evolutionary histories and ecological affinities of plants and animals. Moreover, the hypothetical boundary lines between adjacent biogeographic regions constitute zones where biotas that evolved in different regions overlap with endemic taxa. The relevance of this qualitative aspect of biodiversity lies in the fact that a few countries of the world are located at the meeting ground of adjacent biogeographic provinces with such remarkable assemblages of species. In particular, three countries, Mexico, Indonesia, and China, each include in their territories a pair of the six major biogeographic regions of the world. The Nearctic and Neotropical regions meet in Mexico; the Oriental (or Indomalaysian) and Australoasiatic regions meet in Indonesia, and the Oriental and Palearctic meet in China. The case of the confluence of contrasting faunas of different origin in Indonesia is so striking that it stimulated the famous naturalist Alfred Wallace to develop the discipline of biogeography. Such confluence can be reflected in the fact that on a single tree can be seen monkeys, marsupials, cockatoos, and other birds. Several underlying factors of high biodiversity (see the following) coincide with this qualitative aspect in these three countries, but particularly in Mexico and Indonesia, which further increase their biological richness.

Another aspect of historical and biogeographic nature is the relative magnitude of the coastal/marine contour in relation to land area. Overall, when this ratio is high, some components of biodiversity, including ecological diversity and species richness, tend to be high. Though coastlines and the presence of seas are common to many countries, the ratio is particularly high in islands and some countries that are rich in biodiversity (Table II). Notable examples include Indonesia (an archipelago of several hundred islands), Australia, Madagascar, and Mexico. Table II shows that most of the megadiversity countries have coastline-to-area ratios that are relatively similar, with the exception of the Philippines and Indonesia, two countries constituted by archipelagoes, in which the ratios are particularly high. Mexico, in addition, includes extensive coastlines on two major oceans, the Pacific and Atlantic. Colombia is in a similar situation, although the ratio is lower. Mexico also has a sea of its own, the Sea of Cortez, which has resulted in significant levels of spe-

TABLE II

Presence of Coral Reefs and the Relationship between a Country's Surface Area and Coastline for the Twelve Most Important Megadiversity Countries

Country	Presence of coral reefs	Area (km^2)	Coastline (km)	Ratio of coastline/area
India	Yes	3,287,590	7,000	0.002
Philippines	No?	300,000	36,289	0.120
Madagascar	Yes	587,041	4,828	0.008
Australia	Yes	7,741,220	25,760	0.003
Colombia	Yes	1,138,914	3,208	0.002
Brazil	Yes	8,511,965	7,491	0.0008
Mexico	Yes	1,958,201	11,592.77	0.005
Peru	No	1,285,215.6	2,414	0.001
Venezuela	Yes	916,445	2,800	0.003
Ecuador	Very small	275,800	2,237	0.008
China	No	9,596,961	14,500	0.001
Indonesia	Yes	1,919,442	54,716	0.028

cies richness and endemism, of both marine and terrestrial species. This is why 51% of the marine mammal species of the world, and close to 80% of all sea turtle species, are present in Mexican coastal waters. It is worth noticing that no land-locked country, regardless of its size, is a significant megadiversity country.

D. Ecological Diversity

Ecological diversity is a significant driver of biodiversity, resulting from a variety of factors such as local variation in topography, soil conditions, altitudinal and latitudinal expanse, and overall climatic conditions, as well as historical and biogeographical backgrounds. Several of these factors have been touched upon already, thus only a synthesis and some additional points are necessary here.

Species richness and the generation of endemism tend to be greater where there is significant ecological heterogeneity that allows genetic isolation, local adaptation, and speciation to take place. Such situations can occur because of the existence of a series of isolated mountain peaks, large valleys or drainage systems that become isolated and separated into smaller systems, or areas that, being geologically heterogeneous and complex, produce a variety of soil conditions with well-defined boundaries between them. In addition to the promotion of speciation, these situations may provide the necessary ecological heterogeneity upon which a variety of communities can develop. This is reflected in two important aspects of the definition of megadiversity countries. The first is the occurrence of a rich diversity of vegetation types or ecosystems. The second is the occurrence of a high turnover of species between adjacent localities, leading to a spatially structured diversity of species, known as beta diversity.

Regarding the diversity of ecosystems and vegetation types, Colombia, Mexico, Indonesia, Peru, and India are remarkable in that their territories range from savannas (Colombia) and arid and semiarid ecosystems (the other countries) to evergreen tropical ecosystems, and from coastal vegetation types (e.g., mangroves) to alpine and sub-alpine ecosystems. Brazil, in addition to its great predominance of tropical rain forests, includes other tropical vegetation types such as semiarid cerrado and caatinga.

Regarding species turnover, available data on individual countries are extremely limited, largely because the rate at which the species composition of communities changes spatially depends on the range size of the species and the degree to which species are habitat specialists. However, it can be inferred from the diversity of vegetation types and the magnitude of environmental heterogeneity that the countries that are prolific in these attributes are likely to have high levels of beta diversity. One illustrative case is Mexico, where tropical dry forests expand along a geographic gradient from about 15° to 26°N latitude, in a variety of ecological conditions. At a given locality within this range, the number of arboreal species (with 1.0 cm or greater diameter at breast height) per 0.1 ha is high, with 75 species on average at 20 sites. However, the average proportion of species shared among sites is only 10%. This implies that the turnover of tree species is very high and that the same vegetation type is composed of a series of distinct assemblages of species.

E. Presence of Tropical Rain Forests

Several of the determining factors of species richness and endemism presented here are coincident for tropical rain forests and coral reefs (see Table II). In particular, the presence of tropical rain forests is a significant indicator of the potential for a country to be in the megadiversity group. The remarkable biodiversity of these ecosystems has prompted numerous researchers to investigate the underlying causes of their biodiversity, which include some additional and complementary factors to those described earlier. The most prominent are the following.

1. Climatic Stability

It has been argued that over geological time, tropical regions have had a more stable climate than temperate zones. Thus it is thought that the tropics might have high species richness because of the long time over which species could have accumulated, combined with a lower extinction rate in those stable environments. Such views are being challenged by studies (e.g., of pollen records) that suggest that tropical regions have not been so stable and that, moreover, marked climatic changes have affected the tropics and might have been responsible for the high species richness of some organisms.

2. Pest Pressures

Promoters of the idea that greater pressure from the high diversity of pests in the tropics contributes to species diversity argue that the lack of an unfavorable winter season may promote the persistence of natural enemies (e.g., herbivores and pathogens of plants). The increased pest pressure exerted by these ever-present natural enemies prevents a single, potentially competitively superior species, or group of species, from dominating tropical communities, providing opportunities for numerous species to coexist at relatively low densities. Recent studies confirm that herbivores, for example, are crucial for the maintenance of species diversity in contemporary tropical forests. However, the evidence is limited regarding the role of pest pressure as a factor in generating species richness in these ecosystems.

3. Outcrossing Rates in Plants

The interbreeding with other (often numerous) individuals of the same species, as opposed to self-pollination, is thought to be greater in tropical ecosystems due to the abundance and diversity of tropical pollinators. In addition, the abundance of dioecious plant species (i.e., those having separate male and female individuals, and in which outcrossing is obligatory) is higher in tropical forests. The resulting higher rates of outcrossing may lead to higher levels of genetic variation and thus local adaptation and speciation.

IV. SELECTED MEGADIVERSITY COUNTRIES

The previous description gives some indication of the most likely nations to be considered megadiversity countries. A recent analysis by Mittermeier and Goettsch-Mittermeier (1997) examined 17 countries that collectively possess 66–75% of the planetary biodiversity (terrestrial, marine, and freshwater; Table III). However, not all of these countries comply with the requisites and facets of biodiversity as discussed here. For example, this list includes countries that, although prominent in one or more of the biodiversity facets considered, lack other facets. Some of the selected countries lack tropical rain forest ecosystems, or have a low diversity of ecosystems or cultures, or have relatively low species richness in some groups.

A more synthetic analysis by the same authors considered a shorter list of countries based on a hierarchical scoring of the five most prominent values (scored from 1 to 5) for species richness and degree of endemism of plants, mammals, birds, reptiles, and amphibians. Additionally, even though information is much more limited, consideration was also given to freshwater fishes, butterflies, and tiger beetles (family Cicindelidae). Thus, a country that had the highest value of any of these components was assigned a score of 5, the second highest was given the score of 4, and so on. The total sum of these scores gave a total value that could then be used to compare the different countries. For example, Brazil had the highest values for four components: species richness of plants, mammals, and freshwater fishes, and plant endemism (i.e., 5 + 5 + 5 + 5 = 20; see Table III). It also had three scores of 4 (species richness of amphibians and butterflies and endemism of amphibians, i.e., 12), four scores of 3 (species richness of birds and tiger beetles, and bird and tiger beetle endemism, i.e., 12), one score of 2 (bird endemism), and two scores of 1 (species richness and endemism of reptiles, i.e., 2). The total sum for Brazil is therefore 48. The corresponding values for the most diverse countries are shown in Table IV.

This synthetic analysis confirms the suggestions of the previous sections and can be summarized as follows. Brazil is the most prominent megadiversity country, followed by Indonesia and Colombia. At a considerable distance, Australia and Mexico follow in the fourth and fifth positions, respectively. The relative positions of these remarkable countries change depending on the facet of biodiversity that is considered. On the basis of species richness, the order is Brazil>Colombia>Indonesia>Mexico>Australia. If the emphasis is on endemism, the order is Indonesia>Brazil>Australia, and Madagascar supersedes Colombia>Mexico. The insular character and other attributes of Madagascar highlight its outstanding endemism and place this country in the overall sixth place, close to Mexico. The other countries in Table IV are notable for their species richness (Peru, seventh position; China, eighth position; Ecuador, eleventh position; and Venezuela, twelfth position) or their degree of endemism (Philippines, ninth position). India is remarkable in both species richness and endemism, although it ranks in the tenth position.

All seven of the most prominent megadiversity countries share the following common biological attributes. (1) They have tropical rain forest ecosystems within their territories. (2) They also have marine ecosystems and, to a varying degree, a high coast-to-land ratio (see Table II). (3) Their diversity of ecosystems is considerable and therefore their beta diversity is expected to be high. (4) From a sociological perspective, they share the characteristics of having a very rich diversity of cultures and, with the exception of Australia, of being developing countries and having considerable levels of threat to their biological diversity. Moreover, most of the megadiversity countries listed in Table IV share most of these characteristics. In particular, recent information on species of plants, birds, mammals, amphibians, and reptiles under several categories of extinction risk suggest that no less than 50% of the seriously threatened biological diversity at the global level is concentrated in the megadiversity countries.

TABLE III

Species Richness and Endemic Species (in Parentheses) of Plants and Five Groups of Vertebrates in Seventeen Megadiversity Countries[a]

Country	Plants	Mammals	Birds	Reptiles	Amphibians	Freshwater fish
Brazil	~50,000–56,000 (~16,500–18,500)	524 (131)	1622 (>191)	468 (172)	517 (294)	>3000
Indonesia	~37,000 (14,800–18,500)	515 (201)	1531 (397)	511 (150)	270 (100)	1400
Colombia	45,000–51,000 (15,000–17,000)	456 (28)	1815 (>142)	520 (97)	583 (367)	>1500
Mexico	18,000–30,000 (10,000–15,000)	450 (140)	1050 (125)	717 (368)	284 (169)	468
Australia	15,638 (14,458)	282 (210)	751 (355)	755 (616)	196 (169)	183
Madagascar	11,000–12,000 (8800–9600)	105 (77)	253 (103)	300 (274)	178 (176)	75
China	27,100–30,000 (~10,000)	499 (77)	1244 (99)	387 (133)	274 (175)	1010
Philippines	8000–12,000 (3800–6000)	201 (116)	556 (183)	193 (131)	63 (44)	330
India	>17,000 (7025–7875)	350 (44)	1258 (52)	408 (187)	206 (110)	750
Peru	18,000–20,000 (5356)	344 (46)	1703 (109)	298 (98)	241 (~89)	855
Papua New Guinea	15,000–21,000 (10,500–16,000)	242 (57)	772 (85)	305 (79)	200 (134)	282
Ecuador	17,600–21,100 (4000–5000)	271 (21)	1559 (37)	374 (114)	402 (138)	>44
USA	18,956 (4036)	428 (101)	768 (71)	261 (90)	194 (126)	790
Venezuela	15,000–21,070 (5000–8000)	288 (11)	1360 (45)	293 (57)	204 (76)	1250
Malaysia	15,000 (6500–8000)	286 (27)	738 (11)	268 (68)	158 (57)	600
South Africa	23,420 (16,500)	247 (27)	774 (7)	299 (76)	95 (36)	153
Democratic Republic of Congo	11,000 (3200)	415 (28)	1094 (23)	268 (33)	80 (53)	962

[a] Data from Mittermeier *et al.* (1997).

V. THREATS TO BIODIVERSITY IN MEGADIVERSE COUNTRIES

Megadiversity countries are experiencing significant alterations to their natural ecosystems, and this degradation is a serious threat to their biological resources. The main threats to biodiversity in these countries are the current patterns of land use that lead to deforestation and habitat fragmentation, and species overexploitation. These threats are summarized in the following sections.

A. Deforestation and Habitat Fragmentation

A commonly used measure of habitat destruction is the rate of deforestation. Table V presents compiled statistics on deforestation rates for the period 1980–

TABLE IV

The Ranking of the Twelve Countries with Highest Megadiversity Based on Scores of Species Richness and Endemism (See Text for Details)[a]

Country	Species richness	Endemism	Total
Brazil	30	18	48
Indonesia	18	22	40
Colombia	26	10	36
Australia	5	16	21
Mexico	8	7	15
Madagascar	2	12	14
Peru	9	3	12
China	7	2	9
Philippines	0	8	8
India	4	4	8
Ecuador	5	0	5
Venezuela	3	0	3

[a] From Mittermeier *et al.* (1997).

1989. Rates of deforestation for the 12 megadiversity countries range from 0.4% (Peru) to 2.3% (India and Ecuador) per year. Only four countries have deforestation rates below 1% per year. Even though these figures may seem low, for most countries they imply a significant loss of absolute area per year.

In Mexico, absolute rates of deforestation are on the order of 700,000 ha per year, with by far the greatest amount being tropical wet and dry forests, where most of the biological diversity of the country is concentrated. Moreover, this type of aggregated statistic conceals the fact that particular areas have considerably higher rates of deforestation. Deforestation rates in the tropical forests of southern Veracruz reach values of 4.3% per year. Should such rates continue, tropical forests of the region would be reduced to less than 9% of their original coverage in the first decades of the next century. Such drastic reductions of forest area may drive a significant proportion (about 50%) of the native species to local extinction, according to models derived for the relationship between species and land area on islands. In addition, rates of deforestation do not reflect the fact that the remaining areas of habitat frequently are left in the form of an archipelago of forest islands immersed in a matrix of transformed habitat. Forest fragmentation has been recognized as a significant factor leading to the local extinction of species, reduction of genetic variability, and disruption of several ecosystem processes.

The present rate of worldwide deforestation is close to 50,000 km^2 per year. The loss of species due to habitat destruction may be enormous if this rate is maintained over the next few decades. Botanist Peter Raven estimates that an average of 50,000 species may be lost per year during the next several decades, of which only some 7000 will have been recognized and named.

TABLE V

Demographic, Social, Physical, and Resource Utilization Parameters in Megadiverse Countries[a]

Country	Population (year 2000 estimate, 10^6)	Growth rate, 1993–2000	Rural pop. (% of total in 1993)	Area (ha, 10^6)	Forested area (% of total)	Cultivated area (% of total)	Area under irrigation (% of total)	Firewood use (m^3 (10^3) per year)	Yearly deforestation rate (%, 1980–1989)
Brazil	174.8	1.6	23	851.2	57.3	4.9	6.7	194.27	0.7
Indonesia	212.7	1.5	67	190.5	58.7	9.9	24.3	149.06	0.8
Colombia	37.8	1.5	28	113.9	43.9	3.4	13.5	17.22	1.7
Australia	19.2	1.3	—	771.3	18.8	6	4.6	—	—
Mexico	102.4	1.9	26	195.8	24.9	11.8	26.3	1.3	1.3
Ecuador	12.6	2	43	28.3	55	5.7	34.1	2.3	2.3
Madagascar	17.3	3.2	74	58.7	39.5	4.4	42.1	1.2	1.2
Peru	26.1	1.9	29	128.5	66	2.6	37.6	0.4	0.4
Venezuela	24.2	2.1	8	91.21	32.9	3.5	5.9	0.7	0.7
China	1284.60	1	71	956,100	13.5	9.6	53.6	200.6	—
India	1022	1.8	74	328,759	20.8	50.5	28.9	262.8	2.3
Philippines	74.6	2	48	30	45.3	18.4	28.6	35.98	1.5

[a] Source: UNDP Human Development Report, 1996. N. York.

B. Overexploitation of Species

The overexploitation of species is much more difficult to document and no reliable, compiled statistics are available. Overexploitation of the majority of targeted marine fishes is very well documented, but this generally occurs in the "commons" of international waters. However, some isolated reports indicate that overexploitation is a significant threat to biodiversity in megadiverse countries. For example, in the Brazilian Amazon, studies by zoologist Kent Redford indicate that hunting of vertebrates (mammals and birds) can kill on the order of more than 14 million animals per year. In summary, the great biological richness stored in the megadiverse countries is seriously threatened and urgent measures are needed to deal with this problem, which is occurring on a global scale.

VI. SOCIO-DEMOGRAPHIC CHARACTERISTICS AND THEIR RELATION TO BIODIVERSITY

Of all the issues related to biological diversity, there are very few whose study makes any sense outside of the context of its relation with human populations. Indeed, human populations—their very existence, cultural evolution, size, and distributions—have been and still are intricately dependent and interactive with the biological diversity available to them in the regions where they live. Nowhere are these relations between population, cultures, and biological diversity more evident than in the majority of the so-called "Third World" or "less-developed" nations. Most megadiverse countries share a number of human, cultural, social, and economic traits. These shared characteristics (total human population and growth rates, rural population and its distribution, cultural diversity, land ownership, and levels of scientific development) are discussed here.

Coincidentally, it is in these less-developed regions of the world where the richest biological diversity is present, and where most of the ancient and diverse human cultures and civilizations emerged. These areas presently support the greatest human cultural diversity, represented by the presence of large groups of indigenous peoples. In many megadiverse countries, indigenous groups comprise a high percentage of the total population, commonly ranging from 10% up to nearly 80% in the case of Papua New Guinea.

A measure of the cultural diversity of a country is the number of languages spoken by its people. A total of 6700 languages are spoken in the world, and over half are found in the 15 most biologically diverse countries (e.g., Papua New Guinea alone has over 800 languages). This relationship between biological and cultural diversity is explored by Victor Toledo elsewhere in this Encyclopedia. Usually such a relationship has resulted in cultural traits and land use practices that have increased biological diversity, mostly through the "invention" of new, cultivated species that have allowed the further development of those cultures and others around the world. As illustrated in Fig. 3, countries of both high biological and cultural diversity have usually been centers of origin of crops as well as of agricultural technologies.

A. Rural Populations and Territorial Distribution

Because of the presence of their many different cultural groups, and in stark contrast with most economically developed regions of the world, developing, biodiversity-rich countries have a large proportion of their total population living in rural areas. These groups are highly dependent on natural resources and are usually deeply knowledgeable of their uses and ecological traits. Additionally, in many cases a large proportion of these people live in small villages or hamlets, most of which are distributed within the high biodiversity regions in each country. Because of their close relationship with the surrounding environment and resources, most of these communities are intimately dependent on the conditions of the ecosystems in which they live for their quality of life and the maintenance of their well-being.

B. Land Ownership by Indigenous Groups

The two traits of large populations and predominantly rural living result in patterns of land ownership that may seem "complex" and inefficient to people from developed nations. These patterns are normally the result of well-established traditions and the forms of social organization and consensus building practiced in the communities. Communal land ownership is common, and it can take a variety of forms in different megadiverse countries. Another form of ownership is that of very small landholdings; this system is found in countries like India and China, where its effects are either positive or negative, for the rational use and conservation of biodiversity depending on management practices.

In most cases, indigenous groups own or control

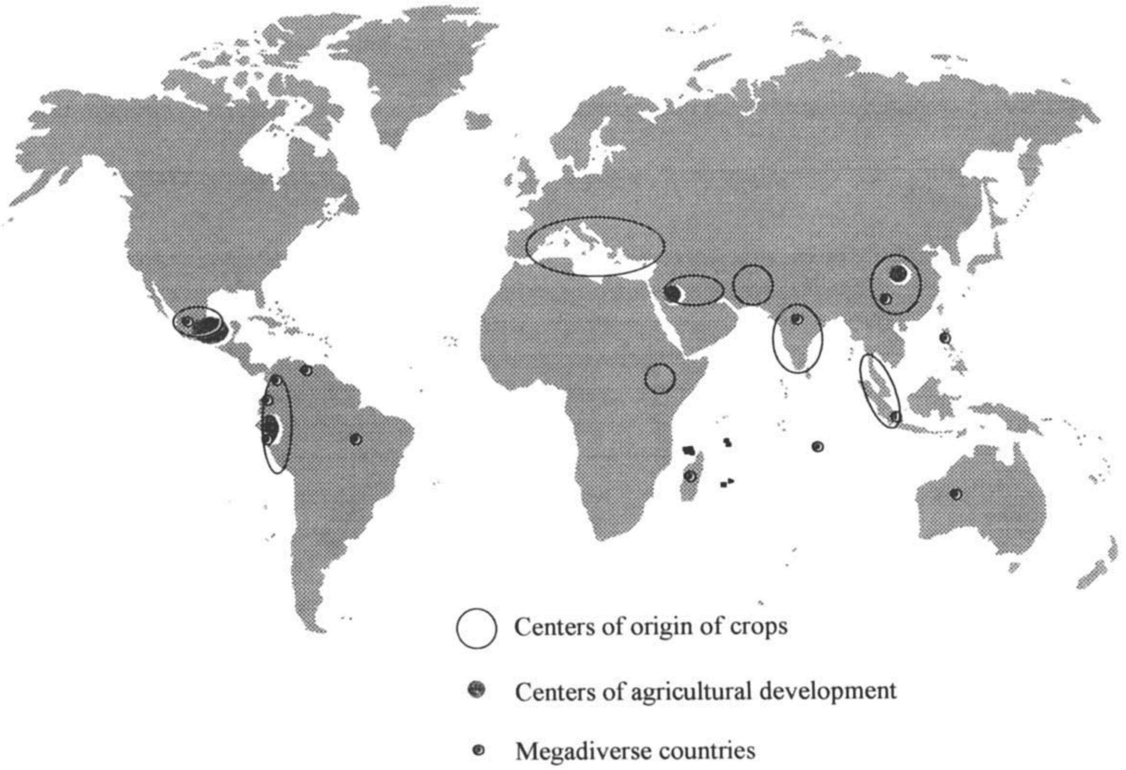


FIGURE 3 Distribution of the twelve megadiverse countries, the centers of origin of crops, and the centers of origin of agriculture. See Table V for the list of countries.

extensive regions of natural forests and other ecosystems in the megadiverse countries. For example, the indigenous groups in the Amazon basin possess over 100 million ha of forested land. In Papua New Guinea, the many cultural groups that live there own close to 97% of the national territory. In Mexico, nearly 75% of all the areas listed as priorities for conservation are under communal ownership, mostly by indigenous or rural groups.

There is an abundant literature describing, with different levels of academic rigor, the benefits and disadvantages of communal land ownership in relation to the rational management and conservation of natural resources, particularly in biologically rich countries. After reviewing much of this literature and visiting the sites and talking to people, it is apparent to the authors that when a strong and healthy system of internal governance exists that allows participation of all interested parties, the community uses and conserves its resources in an admirable fashion. When those internal structures of governance break down for whatever reason, the communal system of land ownership is as destructive of its resources as any other system.

Because of the often large indigenous population in an area, their widespread territorial distribution, and the patterns of land ownership, any program of biodiversity use and conservation in most megadiverse countries must take these stake-holders into account. Such inclusion should not be undertaken as an undesirable and unavoidable necessity, but rather as a wise and intelligent approach that involves local people who are essential to any conservation program. In addition to being the owners of the areas whose resources the conservationists and governments wish to conserve, these indigenous groups normally possess an enormous wealth of valuable and ecologically informed knowledge. Such information is often the result of a long tradition of oral culture that has been passed from one generation to the next, and that makes them truly valuable partners in conservation, management, or restoration programs.

Although a great deal of traditional ecological knowledge exists among the indigenous groups of megadiverse countries, the long-term sustainable use of many of those natural resources remains very uncertain and difficult. This problem is caused by a number of factors, including rising pressure from growing populations to acquire land, increased demands for food production, and the actions of economic interests, often of foreign origin, that are likely to degrade or destroy natural ecosystems and their resident biological diversity.

C. Levels of Scientific Development

Another characteristic that is common to the majority of biodiversity-rich countries (with the sole exception of Australia) is that they have a relatively incipient scientific infrastructure, which in many instances is reflected in the very small number of scientists (taxonomists, botanists, ecologists, etc.) that study their natural resources.

It is arguable to what extent this is the result of the history of colonization—another important characteristic shared by all megadiverse countries. All of these countries have been colonized by a European country for diverse lengths of time, from a few decades to several centuries. In these cases no significant effort was made by the European countries to promote the technical and scientific development of the colonies. Additionally, the subsequent development of these countries was characterized by a neglect of the traditional knowledge on natural resource management of their ethnic communities.

Very often the most important collections of flora and fauna of these countries are found in institutions, herbaria, and museums located in developed countries. This can play a significant role in limiting scientific training and consequently the national capacity to assess and manage natural resources. As described earlier, a great proportion of the planet's terrestrial biodiversity is located in tropical forests in countries that lack the necessary scientific and managerial infrastructure. We believe that this is a major issue that must be addressed when considering the future of the conservation of the world's biological diversity. A detailed discussion of this point is presented in the remainder of this article.

D. A Case Study of Demographics, Cultural Diversity, and Rural Population: Mexico

The following are a few significant statistics for Mexico, but similar patterns are shared by many of the megadiverse countries discussed earlier. These carry high significance in influencing modes of sustainable resource utilization. At present around 6 million people (5 years of age and over) in Mexico have a first language that is not Spanish. This figure has increased between 1930 and 1995 from 2.3 million to 5.6 million persons, although the percentage they represent of the total population has decreased for the same period, from 16% to 6%. They represent at least 54 different ethnic groups, and speak nearly 230 languages. The rural population in Mexico (regardless of language spoken) constitutes about 24% of the total, a considerably lower figure than that of 71% at the beginning of the century; the greatest reduction occurred in the 1950s.

About 90% of all human settlements in Mexico (close to 200,000 registered in the 1995 census) have 1000 or fewer inhabitants (Fig. 4). This proportion is similar to or even higher than that of India, where 73% of all villages have less than 1000 inhabitants. The picture that emerges is clear: the rural population of Mexico is largely atomized, widely dispersed and organized in small villages, with different levels of community organization and various cultural identities. The smallest of these communities are often isolated from most social services and tend to maintain their traditional cultural and social organization traits. Many may succeed in maintaining these traits despite the efforts of external authorities or other sources to exert influence on their internal social organization schemes, or in their modes of production.

Regarding levels of scientific development, Mexico shares the general condition often found in less-developed countries of having a limited scientific infrastructure and little natural resource management experience. However, for different historical reasons, efforts to advance the knowledge of the flora and fauna of Mexico go back to the middle of last century, when plant and animal collections were made, museums were organized, and a limited amount of institutional support was established. With time and the development of public institutions of higher education, many herbaria and museums cropped up in different regions of the country. This brought about the current situation in which taxonomic research, the botanical and zoological survey of the country, and other activities related to an understanding of biodiversity are fairly evenly distributed in many institutions across Mexico. This contrasts with the relatively heavy centralization in the metropolitan area of Mexico City of serious academic activities in other disciplines. The consequences of different levels of scientific development in megadiverse countries are discussed in the following sections when we de-

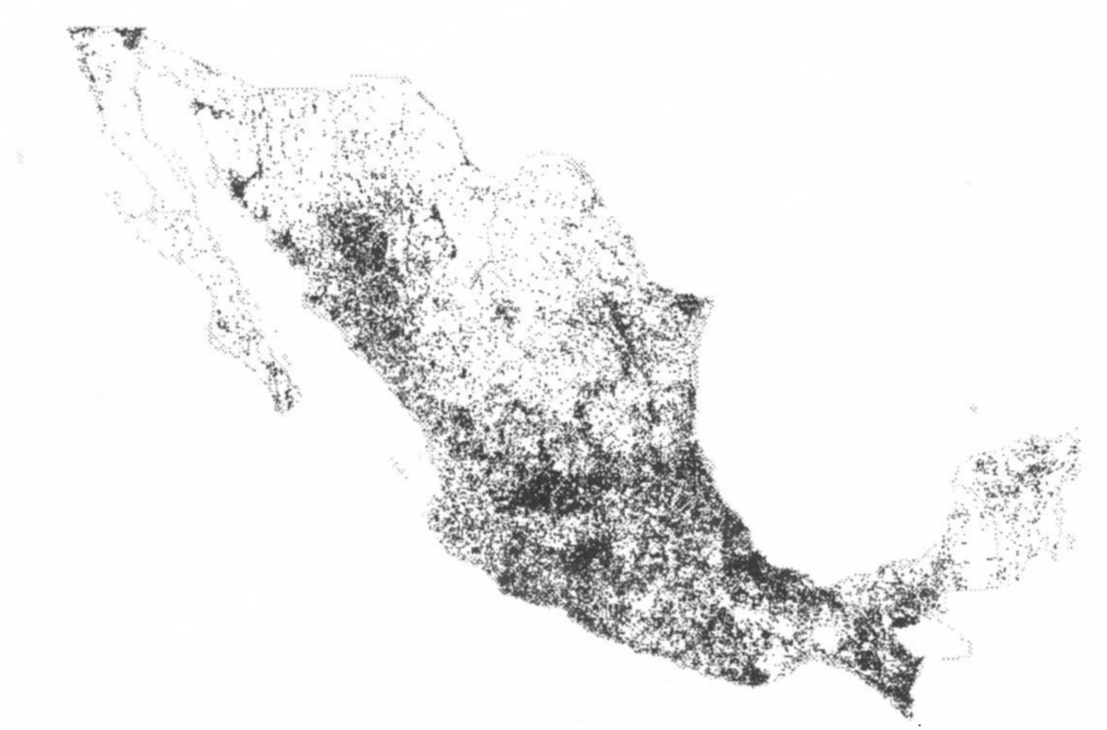


FIGURE 4 A representation of over 55,000 human settlements with 1000 or fewer inhabitants in Mexico. Blank areas in Chiapas and southern Veracruz were not censused. (From the National Institute of Statistics of Mexico, INEGI, 1995, Mexico City.)

scribe the means by which these countries can develop important databases of information about their own biological diversity.

VII. HUMAN AND INSTITUTIONAL CAPACITY FOR BIODIVERSITY KNOWLEDGE AND CONSERVATION

With a few exceptions, a shared characteristic of megadiverse countries is that a very large proportion of the scientific collections that represent their rich biological diversity is deposited in institutions located in diversity-poor but economically rich nations. Most of the vast collections of the Earth's biota (some 3 billion specimens, excluding microorganisms) are located in the largest natural history museums and herbaria of the world. These collections consist largely of specimens collected within biologically rich countries during many decades, or even centuries, by expeditions supported by the institutions of developed countries. Incidentally, perhaps no more than 1% of that vast pool of information can be accessed in automated fashion. Many of these collections represent a heritage of very significant historical value, since they contain biological information of regions that have been seriously altered and that presently have lost much of their original floras and faunas. These natural history museums constitute a vast investment of effort and resources and are an invaluable repository of information.

Much of the reason why these collections are not located in their "resident" countries is because only about 6% of the world's scientists live in the countries that house 80% of the planet's biodiversity. Different estimations coincide in calculating the number of taxonomists in the world to be about 7000. If a similar ratio holds for taxonomists, this means that the megadiverse countries have fewer than 500 taxonomists. This would clearly be an underestimation, since the authors estimate that there must be at least twice as many taxonomists in the megadiverse countries. However, even with this higher figure, this represents a painfully small number of specialists who are unable to adequately study an enormously rich biota.

Comprehensive and scientifically sound biological collections, together with the necessary taxonomic expertise that must go hand-in-hand with the collections, are a fundamental need to advance the ability of a country to know, understand, manage, and preserve its biological heritage. This was recognized at a high official level by the 175 signatories of the Convention on Biodiversity (CBD) during their fourth Conference of the Parties (COP) in Bratislava, Czech Republic, in May of 1998, who concluded that there exists what they called the "taxonomic impediment." This "impediment" results from the severe limitations of institutional and human infrastructure that affect the majority of megadiverse countries, which also happen to be the less economically and scientifically developed. These limitations are illustrated by the very small or nonexistent number of taxonomists, systematists, ecologists, and other specialists trained at a high level, and the absence or very limited number of zoological museums and herbaria that have reliable and long-term institutional support.

Table VI is a compilation of the herbaria (with the number of specimens they contain), botanical gardens (with the number of taxa reported under cultivation), and zoological collections in the 12 most biologically diverse countries. It is clear that the majority have a serious shortage of scientific collections of their floras and faunas, with the notable exception of Australia and to a lesser extent India, China, and Indonesia in Asia and Brazil, Mexico, and Colombia in the Americas.

It is worth noting that for developed countries of the temperate regions the situation is dramatically different. The United Kingdom, France, and the United States, for example, have between 10 and 100 times more herbaria and specimens housed in them than any of the 12 megadiverse countries. The comparison is almost the same for zoological collections and botanical gardens. Obviously, only a fraction of those specimens belongs to their national floras and faunas, which makes them, together with a few other European countries like the Netherlands and Germany, the most important repositories of the biological information of the majority of megadiverse countries.

At the Bratislava COP meeting, an agreement was reached on the imperative need to remove this "impediment" with the aim of fulfilling the objectives set by the CBD within a shorter period of time. However, the training of personnel at a high level, as is required for specialist taxonomists and systematists, and the extensive development of scientific collections (which in turn requires long-term institutional support) are complex, lengthy, and expensive processes. Most of the largest of such institutions in the world are the result of a tradition of many decades, and even centuries, of sustained institutional effort and economic support. This does not mean that countries that do not have, or that need to develop more, such human and physical infrastructures should abandon any attempts to embark themselves in this process. Yet it seems to be far more effective to simultaneously dedicate efforts to develop a strong capacity to acquire, systematize, interpret, and utilize the biological information already existing in the many local and foreign museums and herbaria. Examples of such efforts, although not abundant, demonstrate different strategies in how to bridge the gap of

TABLE VI

Institutional Infrastructure Related to Biodiversity in Megadiverse Countries[a]

Country	Number of herbaria	Accessions in herbaria (10^6)	Number of botanical gardens	Number of taxa in botanical gardens (10^3)	Number of zoological collections
Brazil	88	3.2	23	17	18
Indonesia	6	1.67	5	70	3
Colombia	23	0.66	13	3	8
Australia	38	5.32	63	100	4
Mexico	52	2.3	35	7	25
Madagascar	2	0.08	2	5	2
Peru	11	0.45	6	—	3
China	306	15.5	68	27	7
Philippines	9	0.29	9	17	6
India	52	3.7	68	85	27
Ecuador	10	0.26	2	0.5	3
Venezuela	15	0.63	7	1	15

[a] From Heywood and Watson (1995).

knowledge about the biological diversity of a country. Some of these examples are presented in Section VIII.

VIII. CASE STUDIES OF MANAGING BIOLOGICAL INFORMATION IN MEGADIVERSE COUNTRIES

Most of the examples of development at the national level of the capacity to acquire, systematize, and utilize information about the biota of a country occur, paradoxically, in biodiversity-rich countries. As we have mentioned, by far the majority of the scientific information about the flora, fauna, and microorganisms that live in a region or country is housed in the systematic collections of plants and animals. Such information is complemented to a certain degree by works such as written floras and faunas, updated checklists of organisms, and detailed ecological accounts of groups of plants or animals. However, only the scientific collections contain the wealth of information that covers the historical dimension of which organisms used to live in what now are profoundly disturbed areas. A second dimension, the geographical, is also characteristic of scientific collections, found in the representation of specimens that records the spatial and ecological distribution of a species and its morphological and genetic variance within its range.

Despite the very considerable wealth of information contained in most scientific collections, the access to such information has been essentially restricted to the scientists (taxonomists and systematists) who work on and take care of those collections. However, there has been an intense effort in putting together different catalogs and directories for biodiversity information in several developed countries. These mostly take the form of accounts of collections for diverse organisms, catalogs of described species, lists of endangered species, bibliographic sources, or environmental studies.

A. The Comisión Nacional Para el Estudio y Uso de la Biodiversidad (CONABIO), Mexico

CONABIO was created in March 1992 as an interministerial organization with the mission of coordinating the actions and studies related to the knowledge and preservation of biological diversity, as well as promoting and stimulating scientific research activities for the exploration, study, protection, and utilization of biological resources. Its aims are to preserve Mexico's ecosystems and to develop criteria for their sustainable management. Although CONABIO is placed within the Ministry of the Environment, Natural Resources and Fisheries (SEMARNAP) and receives funding from federal sources, it operates as a private trust fund (Fideicomiso Fondo para la Biodiversidad). This trust fund allows CONABIO to obtain, in addition to federal funds, economic support from contracts, donations, and services, which helps it to run a very efficient operation. The main functions of CONABIO are to:

1. Establish a national program for biological inventories.
2. Synthesize the information relating to biological resources in a permanently updated database.
3. Design and implement a National Biodiversity Information System.
4. Promote projects directed to the actual and potential use of biological resources that are both conventional and nonconventional.
5. Provide advice on technical and scientific aspects related to conservation and management of natural resources to government agencies, nongovernmental organizations, and industry.
6. Produce education materials for the general public, especially for younger people, and provide information about biological diversity and natural resources to all components of Mexican society.

In less than eight years, CONABIO has become the central Mexican institution that compiles data and generates information on biodiversity and natural resources, and in so doing it has been able to bridge the gap between scientific research efforts and the production of useful information that is relevant to policy making in the conservation of natural resources. One of its fundamental activities is the construction of a biodiversity database with the participation of most of the taxonomic community of Mexico. Currently, these data are housed in various institutions, as well as in the foreign herbaria and museums that hold large collections of Mexican plants and animals.

CONABIO does not conduct inventories or fieldwork on its own. Rather, it uses the biodiversity information already stored in national and foreign scientific institutions and funds research carried out by Mexican scientists in the areas that have been identified as priorities in its mandate. An extensive program for the repatriation of data on Mexican species housed in foreign institutions has begun and will continue over the coming years. The biodiversity database is constructed on the basis of the information contained in the label of each specimen included, which is taxonomically validated by experts in the different groups of organisms and then precisely geo-referenced. At present the database includes information on over 2.5 million specimens, and is constructed into a geographic information system (GIS) containing a wide array of data on climate, geology, topography, soils, vegetation types, land use patterns, and other factors. The CONABIO home page is www.conabio.gob.mx.

B. The Environmental Resources Information Network (ERIN), Australia

ERIN is the oldest of all national entities devoted to compiling the biodiversity information of a country and building a database with that information. It has been a model and inspiration for other such systems around the world. The Australian government established it in 1989 as part of the Federal Department of the Environment. As is the case with CONABIO, ERIN's aim is to provide biodiversity and environmental information to assist planning and decision making in the fields of conservation and natural resource management. It also draws information from many sources and institutions and its database is constructed on the basis of specimen information and compiled into a very rich GIS. ERIN is the main source for an array of fairly sophisticated data analyses about the distribution, conservation, and management of the Australian biota. It has become an increasingly powerful analysis tool for many studies of the biogeography and ecology of the country's plant and animal species and the assemblages and ecosystems that they constitute. The ERIN home page is www.erin.gov.au.

C. The Instituto Nacional de Biodiversidad (INBio), Costa Rica

Started in October 1989, INBio has pioneered a substantial effort to survey the biodiversity of Costa Rica. It is a private, nonprofit organization, entitled to receive both governmental and private funds for its operation. Its Board of Directors and the General Assembly include representatives from different sectors of Costa Rican society who are stakeholders in the efforts to conserve the natural resources in the country.

One of INBio's central aims is to carry out a basic national biodiversity inventory. To this end, it has developed an active program of biological surveying of the major groups of organisms and is building scientific collections of the different groups. This effort has involved the deployment of a substantial number of field assistants (mostly laymen and peasants) who are qualified as "parataxonomists" and who do most of the collecting of specimens and the initial taxonomic identification. INBio currently maintains a biodiversity database that provides geo-referenced information (by latitude–longitude), particularly of the specimens that have been collected since it started operating. Another of its stated aims is to support biodiversity conservation and promote new opportunities for sustainable development within the social and economic contexts of Costa Rica.

A third area of INBio activity has been the prospecting for natural products in plants and animals of Costa Rica, which has been an especially active component within its various programs. The organization has a close collaboration with different governmental agencies in charge of the conservation and management of natural resources, and it has served as a constant source of advice and information to them in their programs. INBio's home page is www.inbio.ac.cr.

D. The South African Botanical Diversity Network (SABONET)

The South African Botanical Diversity Network was started in the early 1990s as a program closely linked to the National Botanical Institute (NBI) of South Africa. Its mission is to strengthen the level of botanical expertise, expand and improve herbaria and their collections, and stimulate closer collaborative links among botanists of the 10 countries that form the South African subcontinent (Angola, Botswana, Lesotho, Malawi, Mozambique, Namibia, South Africa, Swaziland, Zambia, and Zimbabwe).

A main objective of SABONET has been the training and support of professional botanists, taxonomists, horticulturists, and plant diversity specialists in these countries. A survey of all scientists working in the fields of botany in the region shows that there are nearly 250 botanists, of which a little over 100 work in institutions outside the area, mainly in Europe. Such personnel should be competent in inventorying, monitoring, and evaluating the botanical diversity of the region in the context of conservation needs and the threats to natural systems imposed by different human activities. To facilitate this, SABONET offers training courses and workshops, organizes botanical expeditions to poorly known areas of the region, and publishes a newsletter. SABONET also publishes a catalog that contains information about all research projects carried out in the region, the botanists in charge of them, a taxonomic index of who is working in the different plant taxa, and a geographical index listing the areas where botanists are working. General inquiries can be addressed to ckw@nbipre.nbiac.za.

IX. CONCLUSIONS

The world's biological riches are unevenly distributed across the continents, and this distribution coincides with the same areas where many of our most ancient and diverse cultures developed. These areas of high biological diversity now face threats from the activities and growing demand for natural resources of today's much larger native populations, but also from demands that originate in distant developed countries. As never before, these resources and ecosystems are under direct and imminent threat. Yet we now have a more complete understanding of these ecologically diverse systems and the impacts of human activities on them on local, regional, and global scales.

Megadiverse countries find themselves pressed between the responsibility, and even the fundamental desire, to preserve their biological diversity and the inescapable need to attend to the living requirements and well-being of their growing populations. The difficulty lies in meeting both the justified demands for development and the responsibility for biodiversity conservation. We believe that in order for these countries to increase the possibilities of having a successful outcome in this "tug-of-war," serious consideration should be given to the following two points.

(1) It is impossible to make informed decisions about the conservation of a country's biological diversity if a scientifically sound platform of information does not exist. Such a platform is built with the training of specialists (taxonomists and systematists), the funding of extensive fieldwork, and the buildup of scientific collections. These are long-term (many decades long) enterprises that should be undertaken in a carefully planned manner. While these efforts are under way, a much faster and efficient way to collect and analyze information is to generate biodiversity databases using new field data and the vast amounts of existing information in herbaria and museums—all incorporated into an accessible geographic information system format. We believe that the examples discussed here offer an array of models that can be followed by other megadiverse countries to enable them to better understand the extent and characteristics of their biological wealth.

(2) In megadiverse countries with large indigenous and/or rural populations, the conservation of biological diversity will require an honest, well-intentioned, and equitable collaboration with the owners of the resources (indigenous people or rural communities) that are the object of conservation or management plans. If this is not done beginning with the initial stages of planning, there will be limited success, if any, in such plans. On the other hand, by carrying out successful collaborative programs using this approach, it will be possible to conserve a nation's biodiversity *and* the cultures, languages, and traditions of the indigenous peoples who own and inhabit the land. Sustained efforts toward achieving this type of partnership for conservation in megadiverse countries will pay off in the long run. Very simply, this "experiment" has seldom been done correctly and, once proven practicable and successful, it could then be adapted and applied to a variety of social, economic, and biological situations.

See Also the Following Articles

BIODIVERSITY AS A COMMODITY • DEFORESTATION • ECONOMIC GROWTH AND THE ENVIRONMENT • INDIGENOUS PEOPLES, BIODIVERSITY AND • SOCIAL AND CULTURAL FACTORS • TROPICAL ECOSYSTEMS

Bibliography

Heywood, V. H., and Watson, R. T. (eds.). (1995). *Global Biodiversity Assessment*. Cambridge University Press, Cambridge, United Kingdom.

Meffe, G. K., and Carroll, C. R. (1997). Principles of Conservation Biology. Second Edition. Sinauer Assoc., Sunderland, Massachusetts.

Mittermeier, R. A., Goettsch-Mittermeier, C., and Robles-Gil, P. (1997). *Megadiversidad: Los paises biológicamente más ricos del mundo*. Agrupación Sierra Madre, S. C., Mexico City.

Sarukhán, J., and Larson, J. (1999). When the commons become less tragic: A study on land tenure, social organization and fair trade in Mexico. In *The Commons Revisited: An Americas Perspective* (J. Burger, R. Norgaard, E. Ostrom, and D. Policansky, eds.). Island Press, Washington, D.C.

Sarukhán, J., and Dirzo, R. (1992). *Mexico Faces the Biodiversity Crisis*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, Mexico City.

World Conservation Monitoring Centre (WCMC). (1992). *Global Biodiversity: Status of the Earth's Living Resources*. Chapman and Hall, London.

BIOGEOCHEMICAL CYCLES

Paul G. Falkowski
Rutgers University

GLOSSARY

anion Negatively charged atom or molecule.

cation Positively charged atom or molecule.

chemoautotroph Capable of growing on inorganic substrates without light energy.

eolian Wind blown.

felsic rock Continental crustal rocks relatively rich in silicon and aluminum.

fixation To make nonvolatile (forming organic molecules from inorganic gases).

heterotroph Growth dependent on the oxidation of organic matter.

mafic rock Oceanic crustal rocks relatively rich in magnesium and iron.

orogenesis Formation of mountains.

photoautotroph Growth dependent on the conversion of light to chemical energy (photosynthetic).

spallation Thermally induced neutron and proton ejection following high-energy proton collision.

VIRTUALLY ALL NATURALLY OCCURRING ELEMENTS are found in living organisms. The biological assimilation, biochemical transformation, physical transport, and geological sequestration and mobilization of many elements are self-perpetuating and self-regenerating, leading to a "cycling" of the elements. Although the bioochemical cycle for each element is unique, general reaction pathways can be discerned depending on the chemistry of the element, its role in metabolism, and its distribution on Earth.

I. INTRODUCTION

Like life itself, biogeochemical cycles are far from thermodynamic equilibrium, have evolved over hundreds of millions of years, and are interdependent, forming biogeochemical systems replete with feedback controls (Schlesinger, 1997). Biogeochemical cycles depend on, and co-evolved with, specific metabolic pathways. Hence, biogeochemical cycles depend on and are a selective force in metabolic (i.e., biological) diversity. Over geological time, biogeochemical cycles are responsible for altering the chemistry of the ocean, atmosphere, and terrestrial ecosystems such that rate-limiting reactions within key cycles modify the tempo and mode of evolution. Here, we examine some of the key biogeochemical cycles in the context of their evolution and biological diversity.

Encyclopedia of Biodiversity, Volume 1

II. THE DISTRIBUTION OF ELEMENTS ON EARTH

A. The Origin of the Elements

The relative abundance of the crustal elements reveals a distribution of much more abundant light elements to lesser abundant heavier elements (Fig. 1). Although the two lightest elements, H and He, were formed approximately 16 billion years ago in the "big bang," all the heavier elements result from fusion of ^{2}He nuclei or fusion/spallation (proton or neutron loss) reactions in stars (Degens, 1989). The fusion reactions involving ^{2}He with itself, H, O, C, or N tend to form even-numbered atomic nuclei, whereas spallation and proton capture leads to odd-numbered nuclei. Additionally, because the nuclei of elements with paired protons are slightly more stable than those with an odd number, there is generally a larger relative abundance of even-numbered elements.

In the origin of our solar system approximately 4.6 billion years ago, elemental composition and planetary accretion were strongly influenced by the gravitational forces of the sun. Planet bodies closer to the sun contain relatively higher proportions of heavy elements than bodies further away. The four innermost planets are approximately three times denser than the outer planets and have solid-rock surfaces that contain a relatively high proportion of metals, especially iron and aluminum. In the accretion process, a further gravitational distillation occurred within the planets. On Earth, the abundant heavier elements, such as nickel and iron, tended to migrate toward the center of the internal gravitational field, whereas lighter elements tended to float above the metal core and, upon cooling, accumulated as a solid surface and crust. The lightest elements formed a gaseous phase. Almost all of the two lightest gases, H_2 and He, escaped the gravitational field and diffused into interplanetary space during this initial period of Earth's history.

The composition of the gases in Earth's atmosphere following accretion is not completely resolved but almost certainly contained high concentrations of CO_2, N_2, and H_2O, HCl, and H_2SO_4. This gas composition is similar to that currently on Venus. Precipitation of minerals and formation of felsic rocks led to the condensation and upward migration of liquid, precipitable water that overlies vast regions of denser, mafic rocks.

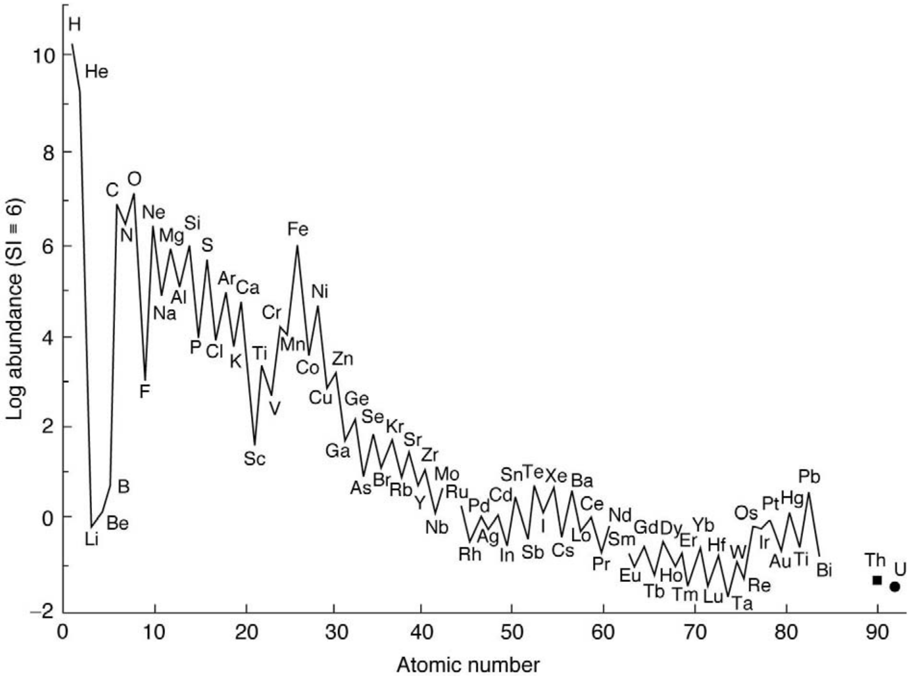


FIGURE 1 Relative abundances of the elements in the universe (based on log abundance of Si = 6). ●, even atomic numbers; ○, odd atomic numbers (reproduced with permission from Williams and Frausto da Silva, 1996, p. 290, *The Natural Selection of the Chemical Elements,* Clarendon Press, Oxford).

Additional water may have been provided by meteoritic bombardment. Based on thermodynamic equilibrium calculations with crustal elements, the acidic gases or hydrated equivalents (e.g., HCl and H_2SO_4) would have solubilized mineral cations in the primordial ocean leading to a seawater dominated by Na^+, K^+, Mg^{2+}, and Ca^{2+}. The anion balance would have been supplied by vulcanism and outgassing from deep crustal sources. The dominant anions were Cl^- and, to a lesser extent, SO_4^{2-} and HCO_3^-. PO_4^{2-} was probably present but to a lesser extent, and fixed inorganic N (as NH_3 or NO_3) was almost certainly very scarce. The pH of the early oceans was probably close to neutral or slightly acidic (6.8–7.0). Additionally, radiogenically produced heat, the high concentrations of greenhouse gases, and geothermal activity would have provided a source of heat to maintain elevated temperatures in Earth's early ocean history.

From these conditions, life arose and evolved. Over the course of Earth's history, there has been a continuous trajectory from the mildly reducing conditions that prevailed at the time of origin of the planet to the highly oxidizing conditions that prevail in the contemporary geological epoch. This oxidation trajectory is driven by biological processes but has also led to the increased diversity of life forms.

B. Phase State Transitions and Elemental Partitioning

The distribution of elements between the atmosphere, lithosphere, and hydrosphere depends, to first order, on the phase state of the element, its chemical reactivity, and its partitioning coefficients between water, the atmosphere, and the lithosphere (Table I). Gases such as N_2 and O_2 partition between the atmosphere and liquid phase according to the solubility coefficient. CO_2, H_2S, and NO_2 undergo hydration to form carbonic, sulfuric, and nitric acids. These acids deprotonate, according to the pH of the aqueous environment, forming anions. Consequently, the total solubility of these species in the aqueous phase exceeds that predicted from the vapor pressure alone, assuming the gases are chemically inert and obey idealized gas laws. Similarly, the distribution of elements between the aqueous phase and the lithosphere is dictated, to first order, by the solubility coefficient of the element, its reactivity or ionic form, and the saturation level. For example, all the alkali and alkali earth metals (i.e., Li, Na, K, Cs, Ca, and Mg) are highly reactive with water and are highly soluble. Precipitation reactions of alkali earth metal cations with carbonate, sulfate, or phosphate anions result in mineral deposits in the solid phase. The movement of elements between the lithosphere, atmosphere, and hydrosphere is mediated by both purely physical/chemical processes and biological processes.

C. Physical/Chemical Transport Processes: Weathering

The flux of elements from continental rock sources to the atmosphere and oceans is promoted by weathering, which is the physical and chemical degradation of rocks to smaller physical pieces and soluble chemical constituents. The physical processes that promote weathering include erosion by water and wind, cracking through freeze–thaw cycles, and active abrasion by the movement of ice (e.g., glaciers), other rock formations (e.g., earthquakes and tectonic uplifting), or rocks (e.g., tidal abrasion). Ultimately, these processes lead to the formation of sediments that are transported by rivers to the ocean. In addition to physical weathering, chemical weathering is promoted by weak acids (e.g., carbonic and sulfuric) in precipitation as well as microbial and plant growth. Over geological time, the weak acids in precipitation remove alkaline cations, such as Mg and Ca, from carbonates and silicates, leading to the mobilization of these elements. Moreover, the production of weak organic acids by lichens, terrestrial plant roots, and the microbial degradation of decaying organic matter greatly promote chemical weathering. Thus, the invasion of land by terrestrial plants led to increased mobility of numerous elements and enhanced the fluxes of materials to the oceans. This results in a feedback between biogeochemical processes on land and in the ocean.

D. The Hydrological Cycle

Weathering is strongly dependent on the hydrological cycle, which is one of the earliest and most critical cycles on Earth. Water is both a solvent and a vehicle for the transport of elements. It is also critically necessary for life. Hence, the distribution of biota is critically related to the availability of water. By far, the oceans are the largest reservoir of water on Earth (Fig. 2). Flux of water into and from the oceans is driven by evaporation and precipitation, processes that are in turn driven by the heat budget of the planet. Over the oceans, evaporation exceeds precipitation; the net difference leads to precipitation over land. Because precipitation exceeds evaporation over land, there is a flux of water from land to the oceans in rivers. The atmospheric

TABLE I

Major Chemical Constituents of the Earth's Crust, Sediments, Ocean Water, and Atmosphere[a]

	Crystal ionic charge and radius		Continental crust		Oceanic crust		Average sediments		Ocean water		Atmosphere	
Element		r (Å)	(wt%)	(Vol.%)	(wt%)	(Vol.%)	(wt%)	(Vol.%)	(wt%)	(Vol.%)	(wt%)	(mol%)
O	−2	1.32	46.40	93.04	43.80	92.57	47.61	91.32	86.0	99.0 as H_2O	23.15	20.95 (O_2)
Si	+4	0.42	28.15	1.04	24.00	0.93	24.40	0.86				
Al	+3	0.51	8.23	0.56	8.76	0.63	6.03	0.40				
Fe	+3	0.64	5.63	0.46	8.56	0.74	3.79	0.30				
	+2	0.74										
Ca	+2	0.99	4.15	1.40	6.72	2.39	7.86	2.54	0.04	0.025		
Na	+1	0.97	2.36	1.31	1.94	1.13	1.36	0.72	1.08	0.11		
Mg	+2	0.66	2.33	0.38	4.5	0.78	2.44	0.39	0.13	0.04		
K	+1	1.33	2.09	1.75	0.83	0.73	2.00	1.61	0.04	0.062		
Ti	+4	0.68	0.54	0.05	0.90	0.09						
Mn			0.095		0.15							
H			0.14		0.2				10.7	See O		
P	+5	0.35	0.105		0.14		0.16	0.003				
S	+6	0.30	0.026		0.025		0.62	0.007	0.09	0.0002		
C	+4	0.16					2.91	0.013	0.28	0.002	0.046	0.03 (CO_2)
Cl	−1	1.81					0.83	1.85	1.94	0.833		
N											75.53	78.09 (N_2)
Ar											1.28	0.93 (Ar)

[a] Adapted from A. Lerman (1979). *Geochemical Processes Water and Sediment Environments*. John Wiley & Sons. New York.

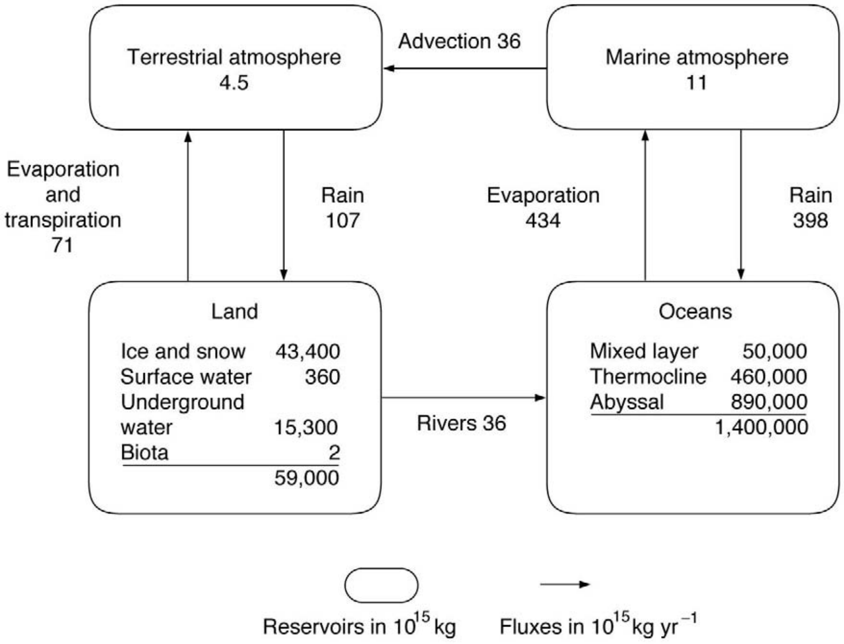


FIGURE 2 Estimates of the global water cycle and its reserves. The accuracy of several of its components is poor, resulting in a closure error for the whole cycle of about a factor of 2. The obvious interactive nature of the cycle makes it impossible to reduce current closure errors without studying the whole cycle. (Reprinted by permission from *Nature,* Chahine, Moustafa T., The hydrological cycle and its influence on climate. Vol. 359, p. 373, copyright 1992, Macmillan Magazines Ltd.)

lifetime of water is extremely short—on the order of ca. 10 days. This lifetime compares with the average for the oceans of ca. 3000 years, whereas deep groundwater can have lifetimes exceeding 10,000 years. These differences in lifetimes are critical determinants of fluxes of elements and climate feedbacks.

E. Vulcanism and Orogenesis

The riverine and eolian (wind-blown) transport of sediments from land to the oceans would, over geological time, erode the continents to the peneplane—the lowest level achievable. Vulcanism and other orogenic processes counterbalance erosion and supply new crustal rock and atmospheric sources of minerals. In the subduction regions of converging tectonic plates, sedimentary rocks are injected back into the deeper crust, where they are reprocessed by heat and pressure, emerging again in uplifting regions and through volcanic eruptions. In addition to the formation of igneous rock, vulcanism provides significant sources of iron, sulfate, and CO_2 to the atmosphere. These elements and materials are transported through the atmosphere and deposited over both the oceans and on land.

F. Biological Transport

Biological processes influence the distribution of elements primarily through phase state transitions. The chemical reduction of inorganic carbon to organic matter is a conversion of a gas to solid phase; the latter is often dissolved in an aqueous phase of the cell cytoplasm. For oxygenic photoautotrophs, the evolution of molecular oxygen leads to the formation of a gas from the liquid phase of water. Biological mediation of phase state transitions is critical for the movement of elements between the atmosphere, lithosphere, and hydrosphere. For example, $CaCO_3$ is by far the largest reservoir of carbon on Earth (Table II). The formation of $CaCO_3$ is the consequence of the biological precipitation by marine organisms and, over geological time, represents an important sink for atmospheric CO_2. The oxidation of organic matter leads to the formation of CO_2, which equilibrates with the atmosphere. Similarly, the biologically catalyzed oxidation of H_2S to S or SO_4 is a phase state transition that influences the partitioning of sulfur between aquatic ecosystems and the atmosphere. It should be noted that the biologically catalyzed phase state transitions need not be direct but can result from

TABLE II

Carbon Pools in the Major Reservoirs on Earth[a]

Pools	Quantity ($\times 10^{15}$ g)
Atmosphere	720
Oceans	38,400
Total inorganic	37,400
Surface layer	670
Deep layer	36,730
Total organic	1,000
Lithosphere	
Sedimentary carbonates	>60,000,000
Kerogens	15,000,000
Terrestrial biosphere (total)	2,000
Living biomass	600–1,000
Dead biomass	1,200
Aquatic biosphere	1–2
Fossil fuels	4,130
Coal	3,510
Oil	230
Gas	140
Other (peat)	250

[a] From Falkowski and Raven (1997, p. 130).

the indirect modification of the redox state of the environment. For example, the oxidation of H_2S occurs spontaneously in the presence of O_2; however, the O_2 is formed from oxygenic photoautotrophs.

G. Biological Assimilation

To accomplish phase state transitions, biological processes catalyze chemical and/or physical reactions. Reaction sequences require an assimilation of the elements. The elemental composition of organisms is dictated, to first order, by the biochemical machinery required for maintenance, growth, mobility, reproduction, and defenses against predation. The four fundamental biochemical constituents that comprise all organisms, namely, proteins, nucleic acids, lipids, and carbohydrates, are primarily synthesized from six light elements: H, C, N, O, P, and S. Of these, N, C, and O are generally reduced, P is oxidized, and S can be both reduced and oxidized. These six elements comprise approximately 98% of all living biomass. The remainder consists of approximately 20 ion-forming elements, including Na, Mg, Si, Cl, Ca, V, C, Mn, Fe, Co, Ni, Cu, Zn, Se, Mo, Sn, and I. The transition metals are frequently used in the mediation of electron transfer (redox) reactions.

III. KEY BIOGEOCHEMICAL REACTIONS AND CYCLES

A. Hydrolysis and Oxidation-Reduction Reactions

Two biologically mediated chemical reactions are key to understanding biogeochemical cycles, namely, hydrolysis and oxidation-reduction processes. Hydrolytic reactions chemically remove from or add water to specific elements or molecules; the reactions are almost always accompanied by the transfer of protons and thus are pH-dependent processes. Hydrolytic reactions can be described in a general form by

$$AH^+ + OH^- \rightarrow A + H_2O$$
$$AOH^- + H^+ \rightarrow A + H_2O$$

Hydration/dehydration reactions can thermodynamically store or discharge energy and determine the reactivity and solubility of elements in the first series (H, Li, Na, etc.) and of phosphorous, (inorganic) carbon, and silica in the environment.

Redox reactions are characterized by the addition or loss of electrons, hydrogen atoms (but not simply a proton), or molecular oxygen. Redox reactions are the primary engines of life, and the consequences of redox chemistry strongly impact the distributions of oxygen, carbon, nitrogen, sulfur, and transition metals in the environment. Redox reactions are always paired; there must be an oxidant and a reductant. The reaction process can be generalized by half-cells:

$$A_{(ox)} + e^- \rightarrow A_{(red)}$$
$$B_{(red)} \rightarrow B_{(ox)} + e^-$$

where the overall reaction is

$$A_{(ox)} + B_{(red)} \leftrightarrow A_{(red)} + B_{(ox)}$$

For each half-cell, the tendency for a substrate to be oxidized or reduced is described by the Nernst equation:

$$E = (E_o + 2.3\ RT)/(nF \log_{10}[A_{ox}]/[A_{red}])$$

where E is the redox potential (in volts), E_o is an arbitrarily accepted standard redox potential, F is the Faraday's constant (= 96,487 C = 1 mol electrons), n is the number of moles of electrons (Faraday's) transferred in the half-cell reaction, R is the Boltzmann gas constant,

T is temperature in Kelvin, and A_{ox} and A_{red} are the activities (or more commonly the concentrations) of the oxidized and reduced forms of the molecules, respectively. The Nernst equation describes the equilibrium condition for both electronic and ionic processes.

To maintain perpetual redox reactions, reductants consumed in one reaction must be regenerated in another (Table III). Hence, by the very fact that redox reactions are paired, a cycle is potentiated. However, to sustain a cycle of redox chemistry requires an input of energy (e.g., solar or thermal) that can be used to drive the initial condition far from thermodynamic equilibrium. Biological reduction of inorganic carbon is one such process.

B. Fundamental Biogeochemical Processes Related to Carbon

The biogeochemistry of carbon is regulated by both hydrolytic and redox reactions. On geological timescales (>1 million years), the concentration of CO_2 in the atmosphere is a balance between crustal outgassing primarily through vulcanism which provides a source of CO_2 to the atmosphere, and chemical weathering of silica-rich rocks on the earth's surface. These reactions can be simplified as

$$CaSiO_3 + CO_2 \underset{\text{metamorphism}}{\overset{\text{chemical weathering}}{\rightleftharpoons}} CaCO_3 + SiO_2$$

TABLE III

Thermodynamic Sequence for Reduction of Inorganic Substances by Hydrogen at pH 7.0 and 25°C[a]

Reaction	E_h(V)	ΔG^b
Reduction of O_2		
$O_2 + 4H^+ + 4e^- \leftrightarrow 2H_2O$	0.812	−29.9
Reduction of NO_3^-		
$2NO_3^- + 6H^+ + 6e^- \leftrightarrow N_2 + 3H_2O$	0.747	−28.4
Reduction of Mn^{4+} to Mn^{2+}		
$MnO_2 + 4H^+ + 2e^- \leftrightarrow Mn^{2+} + 2H_2O$	0.526	−23.3
Reduction of Fe^{3+} to Fe^{2+}		
$Fe(OH)_3 + 3H^+ + e^- \leftrightarrow Fe^{2+} + 3H_2O$	−0.047	−10.1
Reduction of SO_4^{2-} to H_2S		
$SO_4^{2-} + 10H^+ + 8e^- \leftrightarrow H_2S + 4H_2O$	−0.221	−5.9
Reduction of CO_2 to CH_4		
$CO_2 + 8H^+ + 8e^- \leftrightarrow CH_4 + 2H_2O$	−0.244	−5.6

[a] From Schlesinger (1997, p. 234), calculated from Stumm and Morgan (1981, p. 459).

[b] Kcal $mole^{-1}$ per e^-, assuming coupling to the oxidation reaction $1/4\ CH_2O + 1/4\ H_2O \rightarrow 1/4\ CO_2 + H^+ + e^-$ and $\Delta G = -RT \ln (K)$.

The weathering process is stimulated by the hydration of CO_2 in the atmosphere in rain, forming carbonic acid. This is a self-limiting cycle; as CO_2 rises, weathering accelerates leading to a depletion of the atmospheric inventory, which reduces weathering. In the oceans there is an excess of Ca^{2+}, and an equilibrium in the inorganic carbon system is reached. This reaction series can be specified as

$$H_2O + CO_2 \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \rightleftharpoons 2H^+ + CO_3^{2-}$$

and

$$Ca^{2+} + 2HCO_3^- \rightleftharpoons CaCO_{3(\downarrow)} + CO_2 + H_2O$$

Note that the precipitation of $CaCO_3$ leads to the formation of CO_2. In the absence of any biological activity, the geochemical processes would lead to a cycle of inorganic carbon between the three major reservoirs, namely, the atmosphere, oceans, and continents; however, biological processes interact in two critical ways to affect carbon chemistry. First, biological pathways lead to the reduction and oxidation of carbon. Photochemical and chemoautotrophic organisms oxidize alternative substrates to reduce carbon to the equivalent of carbohydrate. These reactions can be broadly classified in the reaction sequence

$$2(H_2A) + CO_2 \rightarrow CH_2O + A_2$$

where A can be sulfur, oxygen, or organic compounds. This is an example of redox reactions applied to the formation of organic material.

Under mildly reducing conditions, photosynthetic carbon fixation is catalyzed by bacteria that couple the oxidation of H_2S or organic compounds to the reduction of CO_2 to form organic compounds. The thermodynamic energy gradient in such systems is relatively low, requiring inputs of ~0.4 eV of photon energy. Such energy demands are satisfied in the infrared spectrum.

C. Biomineralization and its Biogeochemical Inferences

Many organisms precipitate carbonates, phosphates, or silicates to produce minerals. In the marine environment, the precipitation of $CaCO_3$ by phytoplankton, corals, and several genera of marine invertebrates leads to the formation of aragonite and calcite, two forms of carbonate that are relatively resistant to degradation. Virtually all the carbonate precipitated on Earth is the

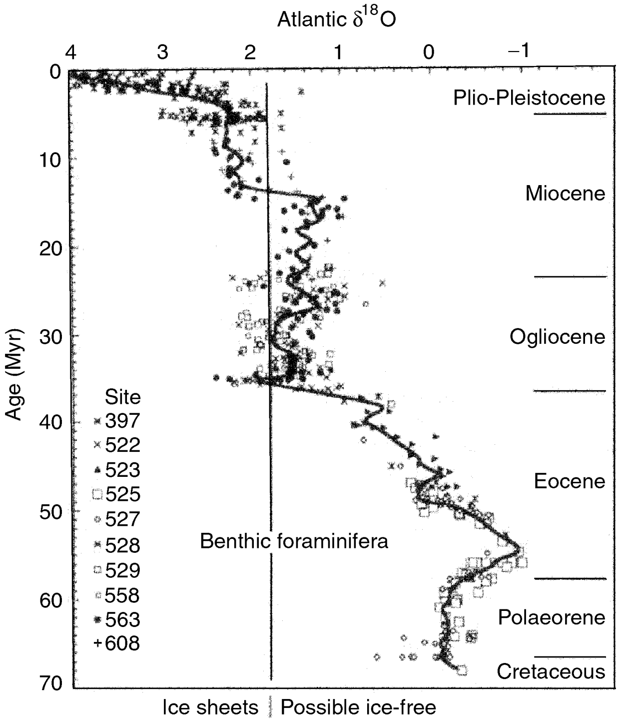


FIGURE 3 Compilation of benthic $\delta^{18}O$ measurements from Deep Sea Drilling Program sites, spanning the past 70 Ma. The long-term increase in $\delta^{18}O$ values reflects cooling of the deep ocean and growth of ice sheets at high latitudes. $\delta^{18}O = [^{18}O/^{16}O)_{\text{standard}} - 1]$, where standard is PDB. (Reprinted by permission from *Nature*, Raymo, M. E., and Ruddiman, W. F. Techtonic forcing of late Cenozoic climate. Vol. 359: 117–122, copyright 1992, Macmillan Magazines Ltd.)

consequence of biological mineralization in the ocean. Over geological time, the production of carbonates produced vast beds of chalk, limestones, and marbles. The deposition of carbonates accounts for the largest reservoir of carbon on Earth. The isotopic fractionation of the stable isotopes of oxygen (^{18}O and ^{16}O) in carbonate sediments is strongly influenced by the temperature of the ocean when and where the carbonates are formed. The isotopic fractionation of the stable isotopes of carbon (^{13}C and ^{12}C) is related to the atmospheric/oceanic inventory of CO_2. Hence, isotopic analyses of carbonates provide clues about both oceanic temperatures and carbon dioxide concentrations over geological time (Fig. 3).

Some groups of marine planktonic organisms, including diatoms and radiolaria, as well as several genera of higher plant grasses, precipitate silica to form hydrated opal. Silica is supplied to the oceans from the weathering of continental rocks and the subsequent flux of soluble silicic acid via rivers. Like carbonates, the precipitation of silicates over geological time leads to sedimentary deposits. In terrestrial ecosystems, several groups of higher plants incorporate silica from soils into stems and shoots.

Under basic conditions, calcium and phosphate ions will spontaneously precipitate to form apatites, which are often mined as a source of phosphate for fertilizers and chemical feedstock. Vertebrates biologically control the precipitation of calcium phosphate to form bones and teeth. Approximately 85% of the phosphate and 95% of the calcium assimilated by vertebrates is incorporated in bones. The fossil record of these biomineral components provides a significant signature of vertebrate evolution but is not a significant component of crustal minerals.

D. The Oxygen Cycle

One of the most important characteristics of Earth's atmosphere is the relatively high concentration of free molecular oxygen (O_2). The mere presence of high concentrations of free molecular oxygen suggests that the chemistry of the planet is far from thermodynamic equilibrium. For all practical purposes, all the oxygen in the atmosphere originated from the photosynthetically mediated oxidation of water by oxygenic photoautotrophs. The oxidation of water to form free oxygen requires an oxidant with a potential of at least 0.83 V at pH 7. The bacterial photosynthetic apparatus could not achieve such a high oxidation potential without evolutionary modification. These modifications included the incorporation of Mn as the transient electron acceptor in the photosynthetic reaction center, the alteration of proteins that permitted the oxidation of water, and the raising of the potential of the primary electron donor by ~0.4 V. The latter was achieved by altering the pigment involved in the photosynthetic reaction from the relatively low energy levels that characterize bacteriochlorophylls to the higher energy levels found in chlorophyll *a*, a pigment that distinguishes cyanobacteria (i.e., blue-green algae) from all other photosynthetic bacteria.

The cyanobacteria are the oldest extant oxygenic photoautotrophs. The exact timing of the origin of cyanobacteria is uncertain, but micropaleological evidence suggests that organisms with features remarkably similar to those of modern cyanobacteria existed in the Archean ocean at least 3.5 billion years before present. Based on geochemical evidence (primarily the oxidation of iron by molecular oxygen), it appears that oxidation of Earth's atmosphere occurred over a relatively short interval of ~100 million years, beginning ~2.2. billion years before present (Fig. 4).

Prior to the photosynthetic production of oxygen, all heterotrophic organisms oxidized organic carbon under anaerobic conditions. The biochemical oxidation of carbon under such conditions is thermodynamically inefficient, often leading to fermentation. For example, in the absence of oxygen, molecules such as acetate are potentially oxidized to methane and CO_2:

$$H_3CCOOH \rightarrow CH_4 + CO_2$$

where the methane is liberated to the environment.

The generation of oxygen led to thermodynamically favorable conditions for the oxidation of organic matter via aerobic respiration. The basic redox sequence for aerobic respiration is

$$(CH_2O)_n + O_2 \rightarrow CO_2 + H_2O.$$

This thermodynamic coupling between the photosynthetic reduction of organic carbon and the oxidation of water is reversed in mitochondria, in which oxygen is used as a sink for electrons and the oxidation of organic carbon is coupled to biochemical energy. Initially, the oxygen could be utilized by single-celled organisms via diffusion; subsequently, however, proteins and structures evolved that facilitated the diffusion and transport of O_2 and led to the evolution of metazoans.

E. Sequestration and Burial

From a biogeochemical perspective, the net evolution (i.e., accumulation) of molecular oxygen in the atmosphere means that oxidation of water and the (bio)geochemical reduction of free oxygen are not balanced, i.e., photosynthetic oxygen evolution must have exceeded oxygen consumption. Given the vast quantities of water on the surface of the planet, oxygenic photosynthesis could produce enormous quantities of oxygen without running out of substrate. When can such conditions occur and what are the oxygen-consuming reactions?

Prior to the net accumulation of oxygen in the atmosphere, oxygen produced by oxygenic photoautotrophs in the oceans would have been exposed to several potential reductants, two of which are especially biogeochemically important, namely, sulfur and iron. After Cl^-,

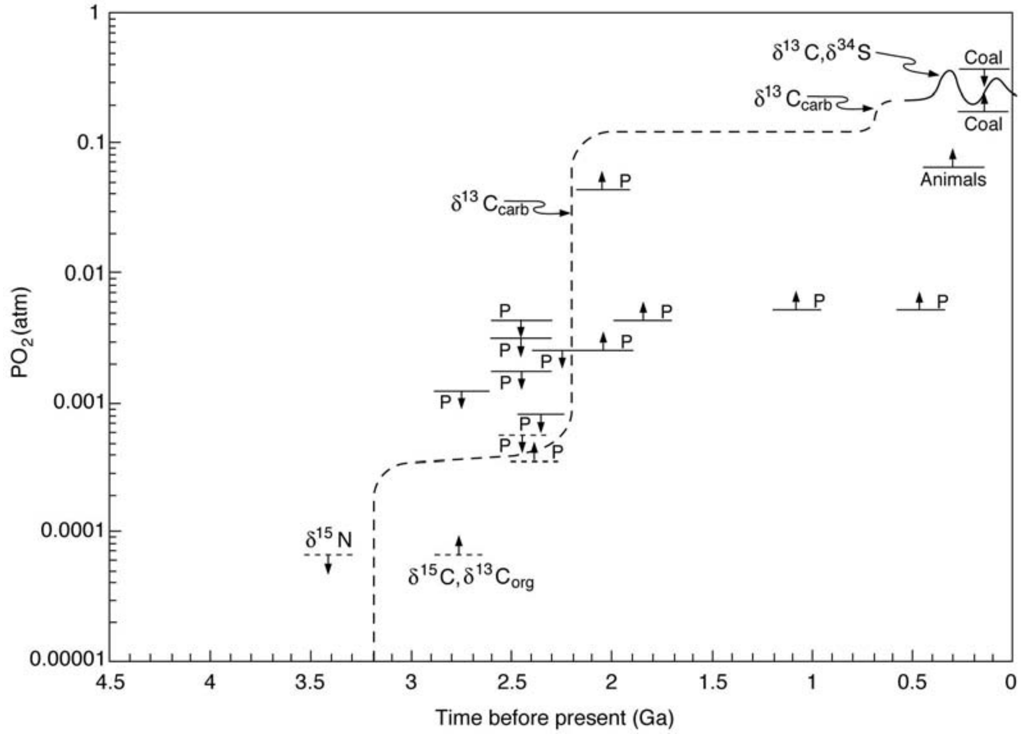


FIGURE 4 The evolution of oxygen on Earth (reproduced with permission from *American Journal of Science,* Rye, R. and H. Holland (1998). Paleosols and the evolution of atmospheric oxygen: a critical review. Vol. 298).

SO_4^{2-} is the most abundant anion in the ocean. Under the mildly reducing conditions of the Archean ocean, dissolved aqueous sulfur in the ocean would have been in equilibrium between a reduced form, probably sulfide (H_2S), and the hydrated, oxidized form H_2SO_4. The latter (sulfuric acid) would equilibrate with the alkaline or earth alkaline elements of the first and second series, resulting in the formation of ionized sulfate salts. The precipitation of sulfate minerals (e.g., gypsum) could occur in shallow seas when evaporation leads to supersaturation and during periods when the influx of Ca from continental sources is high. The latter occurred during periods of rapid invasion of land by terrestrial plants (e.g., during the Triassic Period), where plant growth promoted weathering. Thus, photosynthetic evolution of oxygen led to the formation of vast quantities of sulfate via the oxidation of reduced sulfur anions; sulfur oxidation acted as a sink for molecular oxygen.

Under reducing conditions, iron, as Fe(II), is relatively soluble in seawater, whereas its oxidized counterpart, Fe(III), is virtually insoluble. The photosynthetic oxidation of the Archean ocean not only resulted in the oxidation of sulfur but also led to the oxidation and subsequent precipitation of iron oxides (rust) in ocean sediments. Thus, the biologically catalyzed oxidation of iron served as a second electron sink for molecular oxygen—a sink that ultimately came to represent the largest reservoir for the transition metal on Earth. The deposition of iron oxides would proceed until virtually no soluble iron remained in the ocean. The deposition of iron in Archean sediments is recorded in several deposits, the dating of which provides a record of the oxidation of the atmosphere (Fig. 5).

When the electron sinks in reduced iron and sulfur were consumed, oxygen could be consumed in respiratory reactions involving (photosynthetically) reduced carbon. If, however, there was a net evasion of photosynthetically derived oxygen to the environment (i.e., the oceans and atmosphere), then a stoichiometric amount of organic carbon must have avoided oxidation, i.e., organic carbon must have accumulated. The primary sink for organic carbon is marine sediment in which, in the absence of oxygen, carbon adsorbs to inorganic sedimentary particles and becomes buried in the sedimentation process. The sequestration and burial process occurred over hundreds of millions, if not billions, of years, and it resulted in the accumulation of massive amounts of organic carbon associated with sedimentary and lightly metamorphosed rocks (e.g., shales

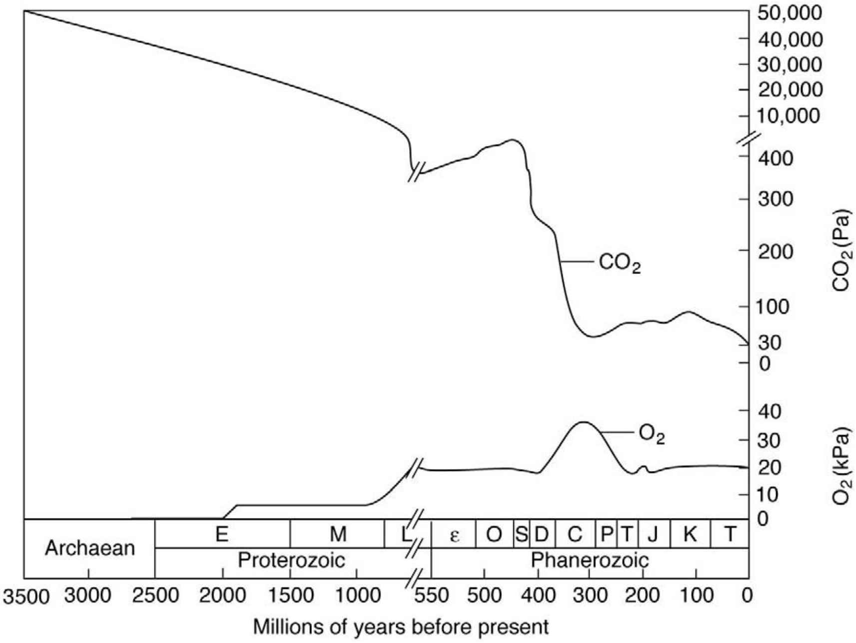


FIGURE 5 A reconstruction of variations in the partial pressures of CO_2 and O_2 in the atmosphere through geological time using data from Berner (1990, 1993) and Berner and Canfield (1989) for the post-Cambrian epochs (i.e., the Phanerozoic). The absolute values and timing for the evolution of oxygen are not constrained in the Proterozoic Epoch. (From Falkowski and Raven, 1997.)

TABLE IV

Annual and Seasonal NPP of the Major Units of the Biosphere, from CASA-VGPM[a]

	Ocean NPP		Land NPP
Seasonal			
April to June	10.9		15.7
July to September	13.0		18.0
October to December	12.3		11.5
January to March	11.3		11.2
Biogeographic			
Oligotrophic	11.0	Tropical rain forests	17.8
Mesotrophic	27.4	Broadleaf deciduous forests	1.5
Eutrophic	9.1	Broadleaf and needleleaf forests	3.1
Macrophytes	1.0	Needleleaf evergreen forests	3.1
		Needleleaf deciduous forest	1.4
		Savannas	16.8
		Perennial grasslands	2.4
		Broadleaf shrubs with bare soil	1.0
		Tundra	0.8
		Desert	0.5
		Cultivation	8.0
Total	48.5		56.4

[a] Ocean data are averages from 1978 to 1983. The land vegetation index is from 1982 to 1990. All values are in petagrams of carbon (1 pg = 10^{15} g). Oligotrophic < 0.1 μg Chl/L; mesotrophic 0.1–1.0 μg Chl/L; eutrophic > 1.0 μg Chl/L. The macrophyte contribution to ocean production is not included in the seasonal totals. (Reprinted with permission from Field *et al.* (1998). "Primary production of the biosphere: integrating terrestrial and oceanic components." *Science* **281**, 237–240).

and cherts). A very small fraction of the buried organic carbon underwent alterations due to heat and pressure (i.e., diagenesis) to become fossilized carbon that literally fuels the industrialized world in the current geological epoch (Table II).

F. Primary Production in the Contemporary World

The carbon remaining in photosynthetic organisms following their own respiratory costs is potentially available for consumption by other organisms. This remaining carbon, or net primary production (NPP), provides the energy requirements of all ecosystems. The distribution of NPP between the terrestrial and marine ecosystems is given in Table IV. These data suggest that terrestrial ecosystems, containing approximately 99% of the plant biomass, account for approximately 55% of the global NPP, whereas marine ecosystems, containing approximately 1% of the plant biomass, account for 45% of the productivity. The huge discrepancy between NPP per unit biomass between these two ecosystems suggests that, on average, the lifetime of marine primary producers is less than 1 week, whereas the lifetime of terrestrial primary producers is on the order of a decade or more. On land, NPP is primarily limited by the availability of water, whereas in the oceans NPP is primarily limited by the availability of essential nutrients, especially fixed forms of nitrogen.

During the past several hundred thousand years, global NPP has remained relatively constant, whereas the distribution of NPP between terrestrial and marine ecosystems has waxed and waned with climatic shifts. Briefly, every 100,000 years or so, changes in Earth's orbital relation to the sun lead to a small reduction in surface temperature, which is amplified (through unknown processes) to produce glacial periods. The advance of ice sheets in the Northern Hemisphere, combined with a reduction in liquid precipitation at low latitudes, leads to the loss of terrestrial NPP by approximately 30%. Simultaneously, however, enhanced upwelling fluxes of nutrients in the oceans combined with a strengthened eolian flux of iron stimulate oceanic

NPP, such that the change between the terrestrial loss and oceanic gain is approximately balanced. During interglacial periods, the situation reverses, with a retreat of the ice sheets and the relaxation of nutrient enrichment of the upper ocean to lead to the situation shown in Table IV for the contemporary world.

G. Oceanic Solubility and Biological "Pumps"

The oceans contain approximately 50-fold more organic carbon than the atmosphere, and on timescales of centuries the former reservoir controls the concentration of CO_2 in the latter. The vertical gradient of inorganic carbon in the oceans is "inverted"—that is, there is a higher concentration of inorganic carbon in the ocean interior than at the surface (Fig. 6). Because the ocean surface tends to equilibrate with the atmosphere, the vertical gradient is maintained by processes other than simple equilibrium diffusion. There are two major processes responsible for the observed gradient. First, cooling of waters at high latitudes increases the solubility of CO_2 while simultaneously increasing the density of the waters. The cold, CO_2-rich waters sink into the ocean interior, where they are transported by density discontinuities and the motion of the earth throughout the world oceans. The round-trip takes approximately 1000 years before the waters are again in contact with the atmosphere in the high latitude from whence they originated the journey. This "solubility" pump is responsible for approximately 50% of the inorganic carbon gradient. The other 50% is a consequence of biological activity. In the upper, lighted portion of the ocean, photosynthetic fixation of inorganic carbon by phytoplankton leads to the formation of organic matter, largely in the form of particulate materials. Gravity, acting on the particles, leads to a persistent "rain" of organic matter into the ocean interior. The organic matter is oxidized by heterotrophic bacteria and other organisms, leading to the production of inorganic carbon. Thus, this vertical flux of organic carbon and its subsequent oxidation effectively pumps up the ocean interior with inorganic carbon. The biological pump is maintained by the upward flux of inorganic nutrients from the ocean interior, especially fixed nitrogen and phosphate. Such nutrients are essential to supporting the photosynthetic activity of phytoplankton in the upper ocean.

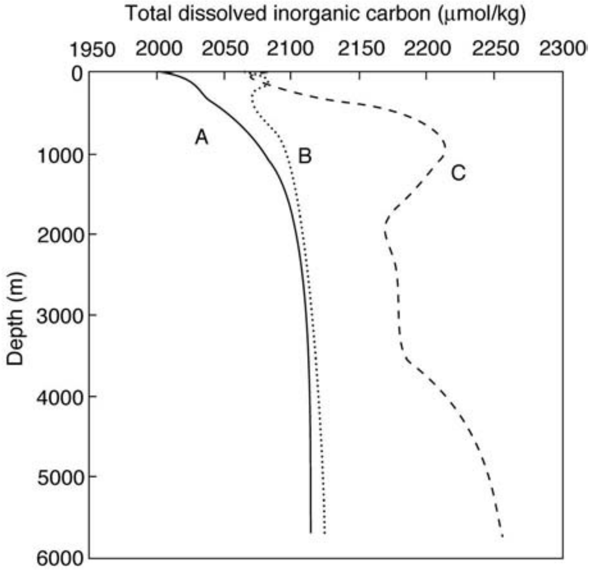


FIGURE 6 Total dissolved inorganic carbon (DIC) concentration in the ocean. (A) Modeled values prior to 1765; (B) modeled values from 1990; (C) observed totals in 1990. The difference between curves C and B represents the contribution of the biological pump to the sequestration of DIC in the ocean interior.

H. The Nitrogen Cycle

All organisms contain nitrogen that is reduced to the equivalent of NH_3 (ammonia). The most abundant form of nitrogen in the environment is N_2, which at the temperature and pressure on Earth's surface is a gas and forms 78% by volume of the planet's atmosphere. Conversion of N_2 to NH_3 requires the addition of three protons and three electrons per atom of N; the balanced equation is

$$N_2 + 6H^+ + 6e^- \rightarrow 2NH_3$$

This reaction is kinetically limited at low pressure and the physiological temperatures, and without some catalytic enhancement there would be no ammonia in the environment. A small subset of bacteria evolved with the capability to "fix" N_2. Nitrogen fixation, or diazotrophy ("eating nitrogen"), is catalyzed by a protein complex (nitrogenase) which is irreversibly inactivated by oxygen. The enzyme complex normally contains an Fe subunit and a Mo, Fe subunit; however, V and Fe can substitute for Mo as mediators of electron transfer in some organisms. Gene sequence analysis suggests that nitrogenase evolved from a single common ancestor, and the enzyme is primarily found in eubacteria, although some Archea, such as methanogens, can also fix nitrogen. *In vivo* nitrogen fixation consumes approximately 16 ATP molecules per mole of N_2 fixed; this high-energy demand requires significant energy supply. In eubacteria, the supply of energy is either light (for N_2 fixing cyanobacteria), or reduced carbon substrates (for heterotrophic bacteria). For methanogens it is H_2. Whatever the energy source, N_2 fixation is a metabolically expensive process, and hence, in the presence of high concentrations of fixed nitrogen, it usually is repressed. Many nitrogen-fixing bacteria live symbiotically with a host organism. One classical symbiosis is found in the Legumacea, which are higher plants such as peas and clover in which nitrogen-fixing bacteria are found in nodules attached to the plant root. In a second symbiosis, some insects, such as termites, contain nitrogen-fixing bacteria within their guts to supply essential amino acids to the host.

In most oxidized environments, fixed inorganic nitrogen is found in the form of nitrate (NO_3^-) or, to a lesser extent, nitrite (NO_2^-). These two oxidized forms of nitrogen are formed from ammonium via the action of nitrifying bacteria. Nitrification is a metabolic process in which the oxidation of ammonium can be coupled to the reduction of inorganic carbon to form organic molecules without light energy. This type of metabolic process is called chemoautotrophy. For example,

$$2NH_3 + 2CO_2 + O_2 \rightarrow 2(CH_2O) + 2NO_2^- + 2H^+$$

The pathway for ammonium oxidation is broken down between organisms that can oxidize the substrate to nitrite and those that further oxidize nitrite to nitrate. Nitrification is a metabolically inefficient process; one mole of NH_3 is oxidized to produce each mole of organic carbon. Hence, the growth rate of nitrifying bacteria is generally low, whereas in the presence of O_2 the rate of ammonium oxidation is generally rapid. Nitrification is confined to a small number of eubacteria; the gene sequence analysis of these organisms suggests the pathways evolved from single ancestral genes and are relatively conserved. Nitrate (NO_3^-) is a thermodynamically stable molecule under the oxidizing conditions of the contemporary world, and as such it is the most prevalent form of fixed nitrogen in the world oceans.

The cycle of nitrogen is completed when oxidized species of nitrogen, especially NO_3^-, are subsequently reduced to form N_2O and N_2. This anaerobic process, called denitrification, is mediated by a wide range of diverse bacteria and Archea. Denitrification is a respiratory pathway in which the terminal electron acceptor is NO_3^- or NO_2^-. Organisms generally reduce these two substrates "opportunistically," that is, when they are deprived of oxygen. The reduction of nitrate to nitrous oxide and N_2 represents a large loss of fixed nitrogen in the environment and largely occurs in anaerobic sediments of marine and lacustrine environments and in areas of soil rich in organic carbon that are periodically flooded (e.g., rice paddies and river deltas).

Although approximately half of all the nitrogen fixed on Earth is due to natural biological activity in the environment, half is due to deliberate fixation by human activities (Fig. 7). Developed in the latter part of the nineteenth century by a German chemist, the Haber reaction permits the formation of ammonium from N_2 by a high-temperature, high-pressure reduction in the presence of H_2. This reaction is the primary source of fertilizer nitrogen for commercial crops, without which agricultural production would be much smaller than it is currently. However, approximately 25% of the applied nitrogen to agricultural crops is lost through denitrification, leading to an increase in N_2O in the atmosphere. A significant fraction of the applied nitrogen in terrestrial ecosystems is further solubilized in waters, flowing either into groundwater or through surface waters to the sea. Most of this fixed nitrogen seldom makes its way to the open oceans; rather, due to the high rates of denitrification on continental margins (especially in sediments on the margins), the fixed nitrogen tends to be lost to the atmosphere through denitrification.

I. Profiles of Elements in the Sea

In the oceans there are three basic profiles to describe the distribution of elements. If an element is conservative, and is either extraordinarily abundant or does not interact with particulate matter (i.e., biological material), it will be homogeneously distributed. The soluble salts in seawater represent a vast quantity of alkaline metals, halides, and sulfates. These elements are so abundant relative to biological demands and biomass that they behave as conservative tracers; that is, their distributions are effectively uniform through the oceans. If an element is biologically reactive and can be absorbed in surface waters and subsequently transported by sinking to the ocean interior, the interior ocean can be enriched in the element relative to the surface. Such elements are said to have "nutrient-like" behavior. Examples of such elements include C, P, fixed N, Si, Fe, Cu, Ni, Cd, Hg, and the rare earth elements (REE's). It is important to note that simply because an element has a nutrient-like behavior it is not necessarily a true nutrient and vice versa. A REE, such as La^{3+}, is highly reactive with sulfhydryl (–SH) groups found in most proteins and hence can be "scavenged" from the surface ocean by particles. In contrast, sulfate is so

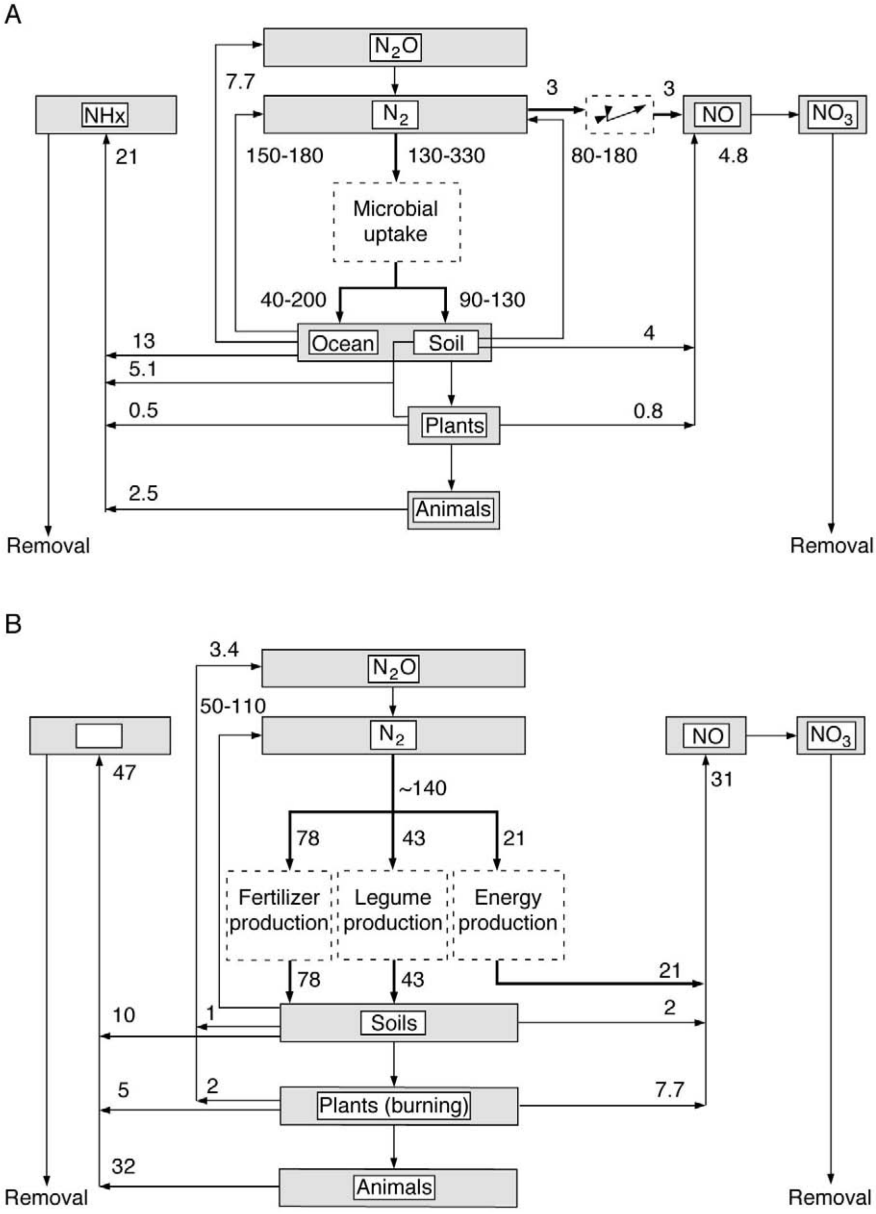


FIGURE 7 Global N-fixation rates and emission to the atmosphere (Tg N year^{-1}) from (A) preindustrial and (B) anthropogenic sources. The boxes with dashed lines represent processes that convert unreactive N to reactive N (reproduced with permission from Gallaway *et al.* (1995) N-fixation, anthropogenic influence. *Global Biogeochem. Cycles* **9**, 235–252).

abundant in the ocean that even though it is used as a nutrient in formation of proteins, the level of utilization is so small in comparison with its concentration that it behaves more like a conservative tracer than a nutrient. The third category of elemental distributions leads to an elevation in surface water and is primarily associated with an upper ocean source. One example is the vertical profile of O_2, which is generated by photosynthetic activity in the upper ocean and consumed in the sea interior by respiration.

J. Sulfur

Sulfur is an integral element in two amino acids, methionine and cysteine, and hence is an essential element for protein synthesis in all organisms. The biogeochemical cycle of sulfur is primarily dependent on microbial metabolism. Under anaerobic conditions, sulfate is used as an electron sink for a wide variety of anaerobic heterotrophic microbes, leading to the formation of H_2S. This same gaseous product can be photochemically oxidized by anaerobic photosynthetic bacteria to produce reductants for inorganic carbon. Thus, a simple cycle of sulfur can be generated under totally anerobic conditions, as long as an external energy source (e.g., light) is available to keep the system far from thermodynamic equilibrium. H_2S is also produced from geochemical reduction of sulfur in the earth's mantle and is emitted to the atmosphere and oceans through vulcanism and geothermal springs (e.g., hyperthermal deep-sea vents). The reduced sulfur flux at deep-sea vents provides a reductant for carbon fixation in symbiotic non-photosynthetic (i.e., chemoautotrophic) bacteria found in the guts of a variety of invertebrates that colonize vent regions. This metabolic pathway is also sustained by photosynthesis, albeit indirectly. The thermodynamic gradient that allows chemoautotrophic sulfur-oxidizing bacteria to extract reducing power from H_2S depends on the oxidation of the ocean by oxygenic photoautotrophs. Under aerobic conditions, sulfur is primarily found in an oxidized form, usually as a sulfate salt, that is assimilated and reduced by photoautotrophs and microbes to a sulfhydryl (-SH). Hence, sulfur is both a source and a sink for metabolic electrons.

In the oceans, sulfate is the second most abundant anion following chloride. The biological assimilation and reduction of sulfur by marine microbes and plankton leads to the formation of volatile sulfur gases, notably dimethylsulfide and, in anaerobic aquatic systems, H_2S. Upon exposure to oxygen in the atmosphere, these reduced forms of sulfur are oxidized and hydrolyzed to form sulfate salts that are transported as aerosol particles. Additional sources of aerosol sulfate include volcanic eruptions and the combustion by-products of fossil fuels, especially coal. Aerosol sulfate plays two important roles in climate feedbacks. First, the particles are hygroscopic, and upon absorption of water vapor they enhance the scattering of shortwave solar radiation (i.e., the planetary albedo). This light-scattering process tends to reduce the heating of the earth by the sun (the so-called "white house effect"). Second, as hygroscopic particles, aerosol sulfate particles provide a nucleation site for water precipitation in clouds, i.e., the particles act as cloud condensation nuclei. The atmospheric lifetime of sulfate particles is short, on the order of weeks. As water vapor builds in clouds, or the clouds are advected aloft and cool, precipitation of liquid water (i.e., rain) washes sulfate from the atmosphere. When aerosol sulfate loading is excessive (e.g., when fossil fuel combustion is high), a small fraction of hydrated sulfate in rain is not buffered by cations, and the sulfate reacts with water to form a dilute solution of sulfuric acid (H_2SO_4), producing acidic rain. Note that the transport of sulfur from continental and aquatic sources to and from the atmosphere is a consequence of phase state transitions, redox reactions, and hydrolysis.

K. Phosphorus Cycle

Phosphorus is an essential component of all cells; the element forms the backbone of nucleic acids, without which cells cannot reproduce. Phosphorus is also incorporated in nucleotides, sugars, proteins, and lipids. In vertebrates, almost all the phosphorus is combined with calcium to form hydoxylapatite, which is the central molecule to build bones.

In the environment, phosphorous is found almost invariably in the oxidized form as hydrated phosphate. Mineral phosphate salts erode and weather from continental sources and are carried to aquatic ecosystems as soluble ions. There is no eolian phosphorus source and no significant transfer of phosphate from aquatic or terrestrial environments to the atmosphere. Unlike carbon, nitrogen, phosphate, or oxygen, phosphorus does not undergo significant biologically mediated redox reactions; the chemistry of phosphorus is solely based on hydrolysis.

L. Trace Metals

Transition metals play a key role in biological electron transfer (redox) reactions and biological processes in turn affect the availability of metals in the environment. The metals are incorporated into protein "scaffolds" in specific orientations and associated with specific ligands such that the electron transfer reactions are highly specified and optimized. The transfer of electrons to given substrates is often keyed to specific metals. For example, electron transfers involving O_2 frequently are mediated by Mn, whereas electron transfers involving N frequently are mediated by Mo. The availability of transition metals, in turn, is dictated by solubility and redox conditions. Under anaerobic conditions, Mn and Mo are often sequestered as sulfide precipitates, but they become more available under oxidizing conditions. In contrast, Fe, which is a required electron transition metal in all organisms, is abundant under reducing conditions but scarce under oxidizing conditions.

As previously noted, electron transfers through iron are essential to both anaerobic and oxygenic photosynthesis, heterotrophic respiration, and nitrogen fixation.

In some cases, anaerobic bacteria can oxidize Fe(II) to Fe(III), thereby precipitating iron as a ferric hydroxide in the sediment. Although iron is the most abundant transition metal in the earth's crust, its concentration in the open ocean is extremely low, averaging no more than ~1 n*M*. To satisfy iron fluxes, eolian sources of iron are critical. Wind-blown dust from major desert areas supplies much of the iron to the open oceans. Consequently, open ocean areas, far removed from continental dust regions, have extremely low iron concentrations that result in a direct limitation of photosynthetic electron transport. Three such areas are the Southern Ocean, the eastern Equatorial Pacific, and the sub-Arctic Pacific. Ice core records suggest that during glacial periods eolian iron fluxes were much greater. The variations in dust supply appear to correspond with changes in the hydrological cycle; warm, wet climate regimes tend to reduce the eolian supply and hence may limit the biological sequestration of carbon in the oceans, whereas cold, dry periods may stimulate oceanic primary production.

Mn is essential for oxygen evolution and, like iron, its concentration is extremely low in the oceans. The Mn requirements for biological activity are lower than those for Fe; consequently, there is no evidence of Mn limitation in either the oceans or the terrestrial ecosystems. Interestingly, Mn and Fe can exchange electrons spontaneously and, in the presence of a redox gradient established in sediments, can cycle between oxidized and reduced sates. Mn has two main oxidation states, namely, Mn(II) and Mn(IV). Organic carbon and NH_4 can be oxidized by Mn(IV), forming CO_2 and N_2, respectively, whereas reduced Mn can be oxidized by O_2 and possibly by NO_3.

Many other transition metals have been co-opted for various biologically mediated electron transfers. Mo is often utilized biologically in electron transfer reactions involving nitrogen. Most nitrogenases contain Mo, as do all nitrate reductases. Copper is frequently found in hydrophyllic proteins that either catalyze single electron transfer reactions (e.g., plastocyanin) or coordinate oxygen binding (in the case of hemocyanin). The availability of these metals is inverse to those of Mn and Fe, namely, under oxidizing conditions, Mo and Cu are more abundant than under reducing conditions (Williams and Frausto da Silva, 1996).

IV. EVOLUTION OF BIOGEOCHEMICAL CYCLES

By definition, biogeochemical cycles are mediated in part by living organisms. The origins of biogeochemical cycles are presumably entwined with the origins of life. Although there is no consensus on the definition of life, to sustain life all living organisms must grow and reproduce. Growth is dependent on the availability of substrates and energy. C, N, P, H, O, and S are required substrates; often these substrates are combined (e.g., all organisms require phosphorous in the form of the phosphate anion, $H_2PO_4^{2-}$). All life forms either oxidize or reduce carbon; hence, carbon is either an electron sink or a source, depending on the thermodynamic basis of the organism's metabolism. Given an external energy source, the redox reactions can run "uphill," where an environmental redox gradient is used to thermodynamically provide the free energy for the subsequent carbon reduction. This metabolic strategy characterizes chemoautotrophs. In photoautotrophs, the energy is externally provided by light. All carbon-oxidizing organisms require an alternative electron sink, such as O_2, NO_3, and SO_4.

All the basic biogeochemical cycles evolved more than 2 billion years ago in unicellular organisms. The evolution of the cycles is interdependent; that is, the requirement by all organisms to obtain reduced carbon for energy and growth created a selection pressure. Under moderately reducing conditions of the early earth, processes that led to N_2 fixation and methane production were more favorable than under oxidizing conditions; hence, these biogeochemical processes are among the earliest. Phylogenetic trees, constructed from 16S RNA sequence homology, suggest that the earliest organisms were nonphotosynthetic, thermophyllic chemoautotrophs that are placed at the root branch between the Archea and Eubacterial kingdoms. These early organisms could have used inorganic substrates, such as H_2, H_2S, and Fe^{2+}, to reduce CO_2 to carbohydrate. Indeed, such organisms persist and thrive in deep-sea vents, in volcanic hot springs deep in Earth's crust, and in other "extreme" environments in which liquid water and suitable oxidizable inorganic substrates are available. Chemoautotrophs are almost certainly the precursors of photosynthetic cells. The evolution of a photosynthetic process in a chemoautotroph forces consideration of both the selective forces responsible and the mechanism of evolution.

Reductants for chemoautotrophs are generally deep in the earth's crust. Vent fluids are produced in magma chambers connected to the athenosphere. Therefore, the supply of vent fluids is virtually unlimited. In the contemporary ocean, the chemical dissequilibria between vent fluids and bulk seawater (which is highly oxidized) provides a sufficient thermodynamic gradient to continuously support chemoautotrophic metabolism. However, the redox gradient in early Earth's oceans did not exist prior to oxygenic photosynthesis. Moreover, magma chambers, vulcanism, and vent fluid fluxes are tied to tectonic subduction regions, which are transient features of Earth's crust and hence only temporary habitats for chemoautotrophs. To colonize new vent regions, the chemoautotrophs would need to have been dispersed throughout the oceans by physical mixing. This same dispersion process would have helped ancestral chemoautotrophs exploit solar energy near the ocean surface. The evolution of photosynthetic organisms led to a highly redundant process; that is, many organisms are capable of reducing inorganic carbon to form organic materials utilizing solar energy. The ecological (i.e., "functional") redundancy ensures a continuity of photosynthetic carbon fixation in all environments in which light and liquid water are available. Currently, there are approximately 20,000 species of aquatic oxygenic photosynthetic organisms belonging to 10 divisions; there are approximately 250,000 species of terrestrial (vascular) plants belonging to 3 divisions (Falkowski and Raven, 1997).

The evolution of oxygenic photosynthesis provided a niche for the co-evolution of ammonium and sulfide-oxidizing bacteria. The action of these organisms led to the chemoautotrophic production of NO_3^- and SO_4^{2-}, which are the most common forms of N and S found in the oceans and most terrestrial ecosystems today. The cycles of these elements are completed under anaerobic conditions, in which the oxidized forms of the elements are used as electron sinks in respiratory pathways. Under many conditions, fixed nitrogen limits primary production in both terrestrial and aquatic ecosystems. In contrast to photosynthesis, N_2 fixation is constrained to a relatively small number of bacteria, and hence the process is "functionally singular." The lack of diversity among N_2 fixers and the subsequent low supply of fixed N to many ecosystems represents a bottleneck in biogeochemical cycles. In the oceans, N_2 fixers appear to be limited by the availability of Fe and P.

See Also the Following Articles

ATMOSPHERIC GASES • CARBON CYCLE • DIVERSITY, ORGANISM LEVEL • GEOLOGIC TIME, HISTORY OF BIODIVERSITY IN • GREENHOUSE EFFECT • NITROGEN AND NITROGEN CYCLE • THERMOPHILES, ORIGIN OF • VENTS

Bibliography

Berner, E. K., and Berner, R. A. (1987). *The Global Water Cycle: Geochemistry and Environment*. Prentice Hall, Englewood Cliffs, N.J.

Berner, R., and Canfield, D. (1989). A model for atmospheric oxygen over Phanerozoic time. *Am. J. Sci.* **289**, 333–361.

Chahine, M. (1992). The hydrological cycle and its influence on climate. *Nature* **359**, 373–380.

Charlson, R. J., and Schwartz, S. E., *et al.* (1992). Climate forcing by anthropogenic aerosols. *Science* **255**, 423–430.

Degens, E. T. (1989). *Perspectives on Biogeochemistry*. Springer-Verlag, New York.

Falkowski, P. (1997). Evolution of the nitrogen cycle and its influence on the biological sequestration of CO_2 in the ocean. *Nature* **387**, 272–275.

Falkowski, P. G., and Raven, J. A. (1997). *Aquatic Photosynthesis*. Blackwell, Oxford.

Fenchel, T., *et al.* (1998). *Bacterial Biogeochemistry: The Ecophysiology of Mineral Cycling*. Academic Press, New York.

Field, C., and Behrenfeld, M., *et al.* (1998). Primary production of the biosphere: Integrating terrestrial and oceanic components. *Science* **281**, 237–240.

Holland, H. D. (1984). *The Chemical Evolution of the Atmosphere and Oceans*. Princeton Univ. Press, Princeton, NJ.

Raymo, M. E., and Ruddiman, W. F. (1992). Tectonic forcing of late Cenozoic climate. *Nature* **359**, 117–122.

Sarmiento, J. L., and Bender, M. (1994). Carbon biogeochemistry and climate change. *Photosyn. Res.* **39**, 209–234.

Schlesinger, W. H. (1997). *Biogeochemistry: An Analysis of Global Change*. Academic Press, New York.

Williams, R., and Frausto da Silva, J. (1996). *The Natural Selection of the Chemical Elements*. Clarendon, Oxford.

BIOGEOGRAPHY, OVERVIEW

Mark V. Lomolino
Oklahoma Biological Survey, Oklahoma Natural Heritage Inventory, and University of Oklahoma

GLOSSARY

biogeography Study of the geographic variation of nature, including variation in any biological characteristics (e.g., body size, population density, or species richness) on a geographic scale.

continental drift Model first proposed by Alfred Wegener that states that the continents were once united and then were displaced over the surface of the globe.

plate tectonics Study of the origin, movement, and destruction of the plates and how these processes have been involved in the evolution of Earth's crust.

Pleistocene Geologic period from 2 million to 10,000 years before the present, which was characterized by alternating periods of glaciation events and global warming.

species composition Types of species that constitute a given community or sample.

species richness Number of species in a given community or sample.

BIOGEOGRAPHY HAS A LONG AND DISTINGUISHED HISTORY, and one inextricably woven into the historical development of evolutionary biology and ecology. Modern biogeography now includes an impressive diversity of patterns, each of which dealing with some aspect of the spatial variation of nature. Given this, few disciplines can be any more relevant to understanding and conserving biological diversity than biogeography.

I. INTRODUCTION

Traditionally, biogeography has been defined as the study of patterns in distributions of geographic ranges (Brown and Gibson, 1983). During the past three decades, however, this field has experienced a great surge in development and sophistication, and with this development the scope of the field has broadened to include an impressive diversity of patterns. Simply put, modern biogeographers now study nearly all aspects of the "geography of nature." Biogeography now includes studies of variation in any biological feature (genetic, morphological, behavioral, physiological, demographic, or ecological) across geographic dimensions such as distance among sites or along gradients of area, elevation, or depth (see Brown and Lomolino, 1998).

A. Fundamentals of Biogeography

Despite the sometimes overwhelming complexity of the natural world, all biogeographic patterns ultimately derive from two very general features of nature. First, as we move along any dimension of the geographic template, environmental conditions tend to vary in a predictable manner. For example, more distant sites tend to be more dissimilar than adjacent sites, environments at higher elevations tend to be cooler and wetter than those at lower elevations, and larger areas tend to capture more solar energy and a greater diversity of environmental conditions than smaller areas. Second, all forms of life differ in their abilities to adapt to geographic variation in their environment. These differences, while including a great diversity of responses (e.g., physiological, behavioral, developmental, and evolutionary), ultimately influence the three fundamental processes of biogeography: *immigration, extinction,* and *evolution.* All the biogeographic patterns we study derive from nonrandom variation in these processes across geographic gradients and across individuals, populations, and species.

B. Early History of the Field

Biogeography has a long and distinguished history, and one inextricably woven into the historical development of evolutionary biology and ecology. The historical development of biogeography had its origins coincident with the Age of World Explorations by Europeans during the eighteenth and nineteenth centuries. Yet the study of geography of nature must be an ancient one. The European explorers were not the first to ask "Where did life come from, and how did it diversify and spread across the earth?" Aristotle asked these same questions, as did many others before and after him, when faced with accounts of strange forms of life from foreign lands.

The development of biogeography into a mature and respected field of science, however, required a much better understanding of variation in what we now call the geographic template and the associated variation in the natural world. It is by no minor coincidence that both evolutionary biology and biogeography developed in earnest during the Age of Exploration. Prior to this time, biologists had "discovered" and described less than 1% of plant and animal forms that we know today. Each new voyage or expedition added to the accumulated information on the earth's environments and lifeforms, and would eventually provide the raw material for the disciplines of evolution and biogeography. These disciplines are interconnected by the knowledge that selective pressures vary across space, and that all lifeforms and their distributions are the product of natural selection.

The early explorers and naturalists did far more than just label and catalog their specimens. They soon, perhaps irresistibly took to the task of comparing their collections across regions, elevations, and other gradients of the geographic template. At the same time, they began to develop explanations for the similarities and differences among the biotas they studied. In fact, most of the persistent themes of the field of biogeography (Table I) were well established during the eighteenth and nineteenth centuries. To be sure, biogeography has made great strides during the twentieth century to become a mature and sophisticated science. It is important to acknowledge, however, that we owe a great deal to the many visionary explorers and naturalists who shared our fascination and asked the same questions about the geographic variation of nature.

1. Historic Explorations of the Eighteenth Century

While motivated to a large degree by the quest for money and power, the Age of European Exploration was also fueled by the call to serve God. It was widely believed that the Creator had placed on this earth a still unknown diversity of organisms—a divine zoo or garden of life. Accordingly, early European explorers believed that there was perhaps no greater way to worship God than to unlock the mysteries of creation. Yet with each new account of some distant biota came information that challenged the prevailing views of creation. Eventually, the growing body of knowledge would overturn the long accepted view that the earth, its climate, and its species were immutable, unchanging in both space and time. More immediately, however, biologists of the eighteenth century were struck by the

TABLE I

Persistent Themes of Biogeography[a]

1. Comparing and classifying geographic regions based on their biotas.
2. Reconstructing the historical development of biotas, including their origin, spread, and diversification.
3. Explaining the differences in numbers as well as types of species among geographic areas.
4. Explaining geographic variation in the characteristics of individuals and populations of closely related species, including trends in morphology, behavior, and demography.

[a] After Brown and Lomolino (1998).

astounding diversity of species. Such diversity presented them with two serious problems, one practical and the other conceptual. First, biologists urgently needed a systematic and generally accepted scheme for classifying the burgeoning wealth of specimens, one that would reflect the similarities and differences among the species. Second, it quickly became clear that there were just too many species to be carried by Noah's Ark. How could all the forms of life, now adapted to many distant and distinct regions across the globe, have originated and then spread from that one landing point?

Carolus Linnaeus (1707–1778), certainly one of the most prominent biologists of all time, took on both of these challenges. In fact, his system of binomial nomenclature is the system we continue to use today to classify organisms. Linnaeus also attempted to rectify the Biblical doctrine with what he and his contemporaries knew about the diversity and geography of nature. This was especially challenging because, like most of his colleagues, Linnaeus was sure that species were immutable. Given this, how could species adapted to a single site and climate (Noah's landing) have spread and become perfectly adapted to a suite of different environments (e.g., alpine tundra, coniferous forests, lowland forests, and grasslands)? Linnaeus' answer: Noah's landing had occurred along the slopes of Mount Ararat, a high mountain near the border of Turkey and Armenia. This mountain is so tall (reaching 16,853 ft above sea level) that along its slopes could be found a succession of environments and communities ranging from subtropical grasslands at the lower elevations to forests and alpine tundra at its summit. According to Linnaeus' hypothesis, each elevational zone harbored a distinct assemblage of animals, each immutable but perfectly adapted to their local environment. When the Flood finally receded, these animals then dispersed to eventually colonize their respective environments across the globe.

One of the foremost challenges to Linnaeus' views came from his contemporary Comte de Buffon (1707–1788), who believed that not only were climates mutable, but species were as well. How else could animals disperse across what are now inhospitable habitats to occupy their present ranges in such isolated regions of the globe? Buffon's theory of the origin and spread of life stemmed largely from his studies of living and fossil mammals, especially those of the Old and New World tropics. He was the first to realize that different regions of the globe, even those with the same environmental conditions, had distinct biotas. This observation was so fundamental that it eventually became biogeography's first law: Buffon's Law.

Like Linnaeus, Buffon also concluded that there was one "landing point," one site where all animals originated. However, this site, or region, was located far to the north of Mount Ararat, somewhere in the Arctic Circle where the early animals and their descendants could gain ready access to both the Old and New Worlds. This is where these life-forms survived the Flood during some earlier period when the earth's climate was much warmer, warm enough such that tropical environments could extend far poleward. Once the floods receded, animals spread southward into the continents and began to diverge in form as they became increasingly isolated on different landmasses.

Other biologists of the eighteenth century, including Joseph Banks and Johan Reinhold Forster, both of whom served as naturalist on voyages of Captain James Cook, were quick to confirm the generality of Buffon's Law: environmentally similar but isolated regions have distinct assemblages of plants and animals. Forster also discussed the relationship between regional floras and environmental conditions and, in turn, between plant and animal associations: two cornerstones of the field now known as ecology. Forster was also one of the first scientists to report that plant diversity increases as we move toward the equator, that islands have fewer plants than the mainland, and that the diversity of insular plants increases with island size and available resources. Later in the eighteenth century, Karl Wildenow (1765–1812) and one of his students, Alexander von Humboldt (1769–1859), confirmed and further generalized both Buffon's Law and Forster's. Toward the end of the century, Augustin P. de Candolle added the important insight that, not only is the distribution of organisms influenced by geographic variation in environments, but they also compete for limiting resources such as food, light, and water.

Therefore, by the beginning of the nineteenth century, biogeographers already had their first "law," they described and tested the generality of a number of related patterns about the geography of nature, and they offered some testable theories regarding those patterns. They were actively working on at least three of the four persistent themes of biogeography (Themes 1–3 in Table I). A number of biogeographers had abandoned the notion that species and climates were immutable. But for the field to advance and become a mature science, two additional, fundamental insights were needed. First, it required a mechanism for the mutability and adaptation of species. As many of us realize, Candolle's observations about competition and the struggle for existence were central to the development of the theory of evolution by natural selection.

These advances were to come in the latter part of the nineteenth century. Second, scientists had to recognize that the geographic template (i.e., the foundation for all of these patterns) also was mutable. That is, the size and relative positions of the continents and ocean basins have changed throughout the history of our planet.

2. Advances of the Nineteenth Century

Many of the most fundamental advances in biogeography, and evolutionary biology as well, have required advances in geology. Until the nineteenth century, the age of the earth was typically assumed to be just a few thousand years, way too brief to allow what we now know to be the requisite time for evolution of its plates and the species that have rafted on those plates. The collective work of nineteenth-century paleobiologists would push the age of the earth back hundreds of thousands and eventually millions of years before the present. Legendary geologists and paleobiologists such as George Lyell (1797–1875) and Adolphe Brongniart (1801–1876), through their studies of fossils, provided incontrovertible evidence for extinction and for changes in regional climate and the elevation of land. How else could they explain the existence of fossils that have no contemporary forms, of fossils from tropical speies found in regions that are now temperate, and of shells and other marine fossils on present-day mountains? A theory of floating and drifting continents (now known as plate tectonics) would await discoveries of twentieth-century marine geologists, but their nineteenth-century forerunners understood that the earth was very old

FIGURE 1 Alfred Wallace's (1876) scheme of biogeographic regions, which attempts to divide the landmasses into classes reflecting affinities and differences among terrestrial biotas. The regions shown are still widely accepted today. Numbers identify subregions. (From Wallace, 1876.)

indeed, and it was mutable. Furthermore, if species (and many thousands of them) went extinct, then there had to be multiple periods of creation (or evolution) to compensate for those losses.

Again, these views of a mutable earth, mutable climate, and mutable species were essential for those attempting to classify biogeographic regions based on their respective assemblage of species (Theme 1), to reconstruct the origin, spread, and diversification of life (Theme 2), and to explain differences in numbers and types of species among geographic regions (Theme 3). However, well into the middle of the nineteenth century many scientists held stubbornly to the idea that not only were species immutable, but so were their distributions. Perhaps the most distinguished champion of this static view of biogeography was Louis Agassiz (1807–1873), who argued that species remain essentially unchanged at or near their sites of creation.

The static view was eventually overturned by the passionate and persuasive arguments of no one less than Charles Darwin. Not only did he propose a general theory for the diversification and adaptation of biotas (i.e., natural selection), but he was one of the world's first and foremost dispersalists and champions of dynamic biogeography. Through his observations during his circumnavigation of the globe on the *HMS Beagle* (1831–1836), his later experiments on dispersal of seeds by animals, and his general synthesis on the origin and distribution of life, Darwin convinced many of his colleagues that long-distance dispersal could account for many of the otherwise perplexing patterns of biogeography. Once he was joined by the likes of Asa Gray and Alfred Russell Wallace, Darwin and his colleagues were able to pull off a major paradigm shift in the field—from the static view of the earth and its species to the dynamic view of biogeography.

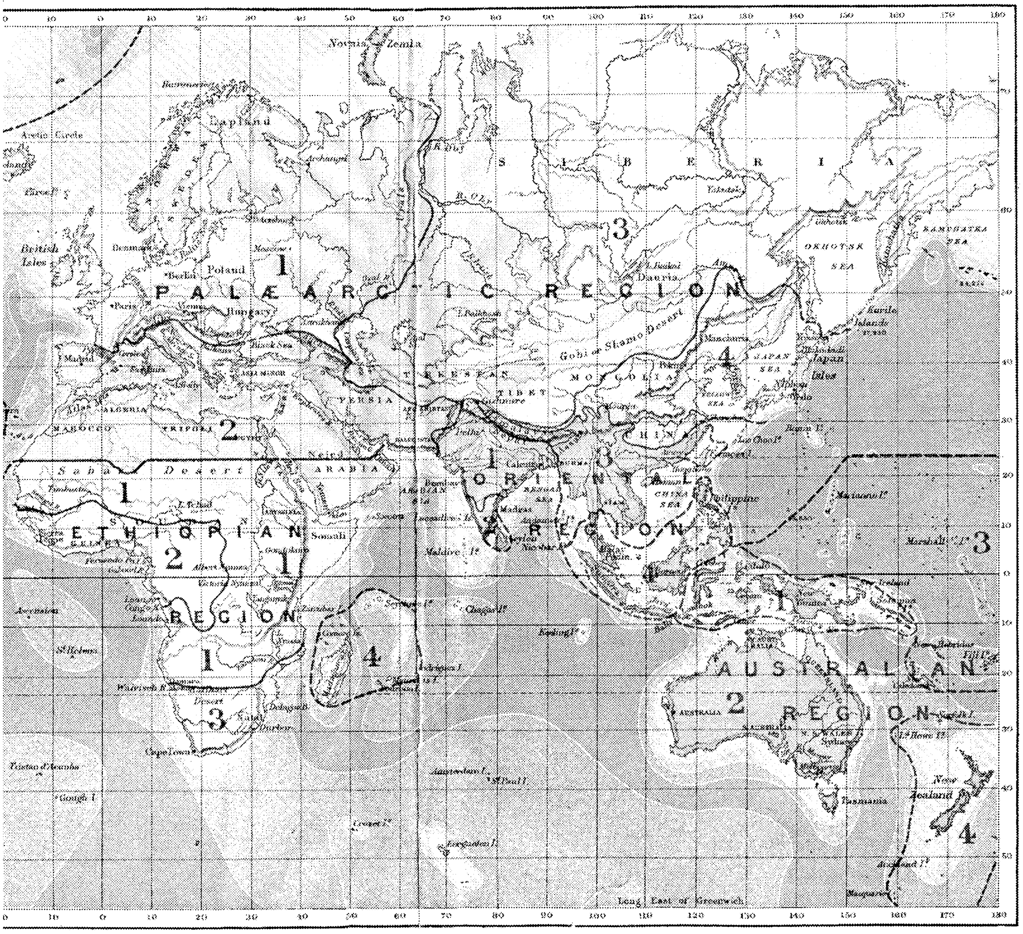


FIGURE 1 (*continued*)

Yet among all of these visionary scientists it is Wallace who is generally recognized as the father of zoogeography, and biogeography in general. While Darwin argued passionately regarding long-distance dispersal (even to the point of soundly criticizing his mentor, Charles Lyell), most of his energies were devoted toward developing and substantiating his theory of natural selection. On the other hand, biogeography was Wallace's life's work. Brown and Lomolino (1998) listed 17 tenets of the field that were developed by Wallace and included in his seminal monographs *The Malay Archipelago* (published in 1869 and dedicated to Darwin), *The Geographic Distribution of Animals* (1876), and *Island Life* (1880). Five of these tenets of biogeography are listed here:

1. Climate has a strong effect on the taxonomic similarity between two regions, but the relationship is not always linear.
2. The present biota of an area is strongly influenced by the last series of geological and climatic events.
3. Competition, predation, and other biotic factors play determining roles in the distribution, dispersal, and extinction of animals and plants.
4. When two large landmasses are united after a long period of separation, extinctions may occur because many organisms will encounter new competitors.
5. To analyze the biota of any particular region, one must determine the distributions of its organisms beyond that region as well as the distributions of their closest relatives.

Using the latter approach and information provided by over a century of naturalists, Wallace developed a scheme of biogeographic regions (Fig. 1) that accurately reflected the similarities and differences among biotas. This same scheme, largely unchanged, is still used today.

For obvious reasons, exploration and biogeographic study of the marine realm have always lagged far behind that of terrestrial systems. Yet by the middle of the nineteenth century, biogeographers had made some significant strides in studying this new frontier. Charles Lyell discussed patterns of distribution of marine algae in his seminal work *Principles of Geology,* first published in 1830. Edward Forbes wrote the first comprehensive monograph on marine biogeography in 1856, in which he divided the marine realm into zoogeographic regions based on latitude, depth, and animal assemblages. In 1897 the great British ornithologist and biogeographer Philip Sclater, who produced a predecessor to Wallace's biogeographic scheme, also developed a scheme for the marine realm based on distributions of marine mammals. Following the lead of earlier biogeographers and also based on his own extensive field studies in southwestern North America, C. Hart Merriam (1894) developed a system of what he termed "life zones" that confirmed earlier observations that elevational changes in vegetation were equivalent to those along latitudinal gradients.

Finally, the countless specimens collected during the late eighteenth and early nineteenth centuries enabled others to begin to analyze geographic variation in characteristics of individuals and populations (Theme 4). C. L. Gloger reported in 1833 that, within a species, individuals from more humid habitats tend to be darker than those from drier habitats (Gloger's Rule). C. Bergmann (1847) found that in birds and mammals, populations from cooler environments tended to have larger bodies than those from warmer environments (Bergmann's Rule). Also, J. A. Allen reported in 1878 that birds and mammals inhabiting cooler environments also tend to have shorter appendages (Allens' Rule).

II. BIOGEOGRAPHY IN THE TWENTIETH CENTURY

A. Dynamics of the Geographic Template

Even the earliest human explorers appreciated the fact that abiotic conditions vary as one moves from one point on the globe to another. On land, precipitation, temperature, seasonality, prevailing winds, soil conditions, and a host of other important factors vary as we move along transects of latitude, longitude, or altitude. Similarly, in the aquatic realm, temperature, currents, pressure, solar radiation, and concentrations of oxygen and dissolved nutrients vary markedly within and among ecosystems. Together, the variation in all of these environmental characteristics combine to form the geographic template, which influences all biogeographic patterns.

Although a complete understanding of all aspects of the geographic template may be a daunting and truly impossible challenge, at a regional to global scale, geographic variation in environmental conditions is quite regular and interpretable. On land, climatic conditions vary in an orderly manner with latitude, elevation, and proximity to mountain ranges or oceans (Fig. 2). Major soil types (Fig. 3) also vary in a similar fashion, partially because soil development is strongly influenced by local climatic conditions, especially precipitation and tem-

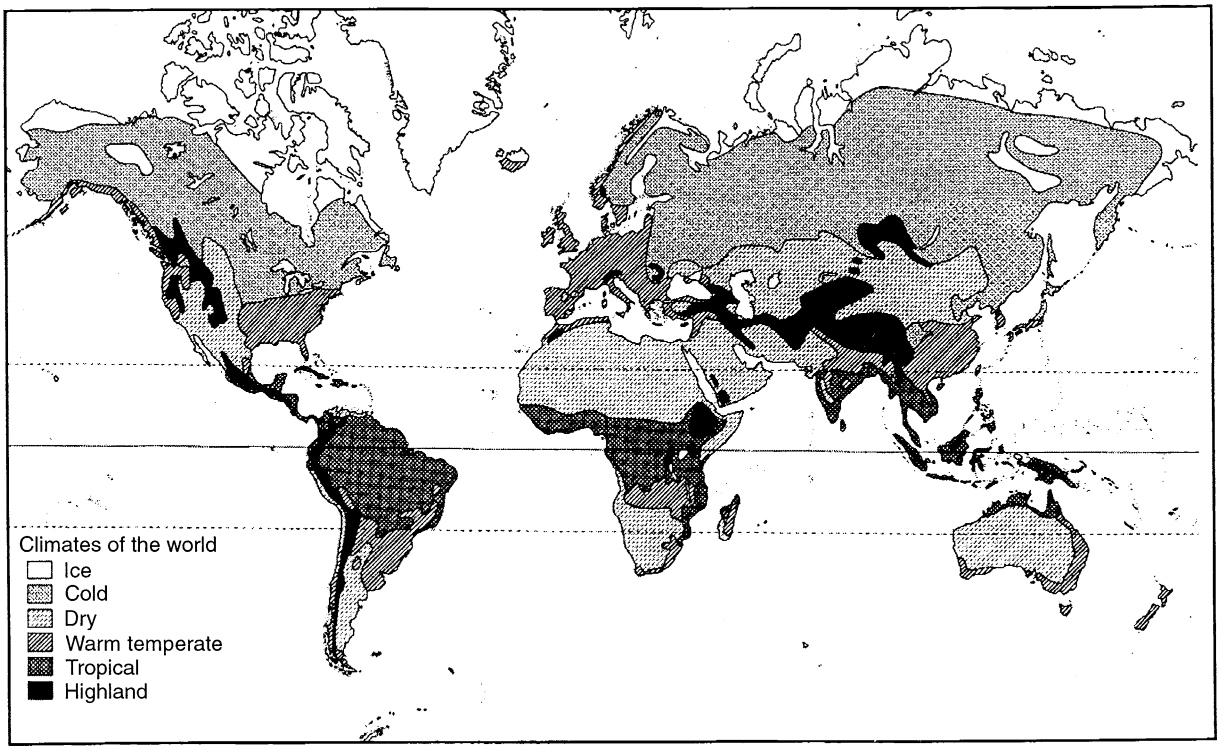


FIGURE 2 Major climatic regions of the world. Note that these regions occur in distinct patterns with respect to latitude and the positions of continents, oceans, and mountain ranges. (After Strahler, 1973.)

perature. In the aquatic realm, although the great volume of ocean waters tends to buffer variation in temperature, surface waters still exhibit a latitudinal gradient in temperature (Fig. 4). In addition, throughout most of the world's oceans light availability and water temperatures tend to decrease while pressure increases with increasing depth.

As we shall see in later sections, such regular variation in environmental characteristics translates into nonrandom variation in biogeographic patterns of organisms, with each one adapted to slightly different environmental conditions. Such adaptations are, of course, the product of a long and complex evolutionary history: a series of innumerable interactions between organisms and their environments. With each successive generation, the abilities of descendants to respond and adapt to local environmental conditions change. Evolutionary change, however, is part of a never-ending battle because environmental conditions include other species, which are also evolving. Just as important, the geographic template has evolved throughout earth's 4.5-billion-year history. Because species distributions and other aspects of their geographic variation are influenced by their interaction with the geographic template, a thorough understanding of its dynamics in space and time is essential if we are to understand any major biogeographic patterns.

By the opening of the twentieth century, each of the four persistent themes of biogeography was well established. Explanations for the major biogeographic patterns could now draw on insights from the rapidly growing field of evolutionary biology, as well as our knowledge of the other two fundamental biogeographic processes—immigration and extinction. In addition, biogeographers of the early twentieth century could tap a great wealth of information on geographic variation of biotas and of the environments that they inhabited. Obviously, a thorough knowledge of this underlying geographic template was essential for understanding patterns in distribution and variation among regions and isolated ecosystems. Yet to develop a more accurate and more comprehensive understanding of the major patterns and processes of biogeography, another major scientific revolution was required.

Biogeographers and most other natural scientists knew a great deal about contemporary environments,

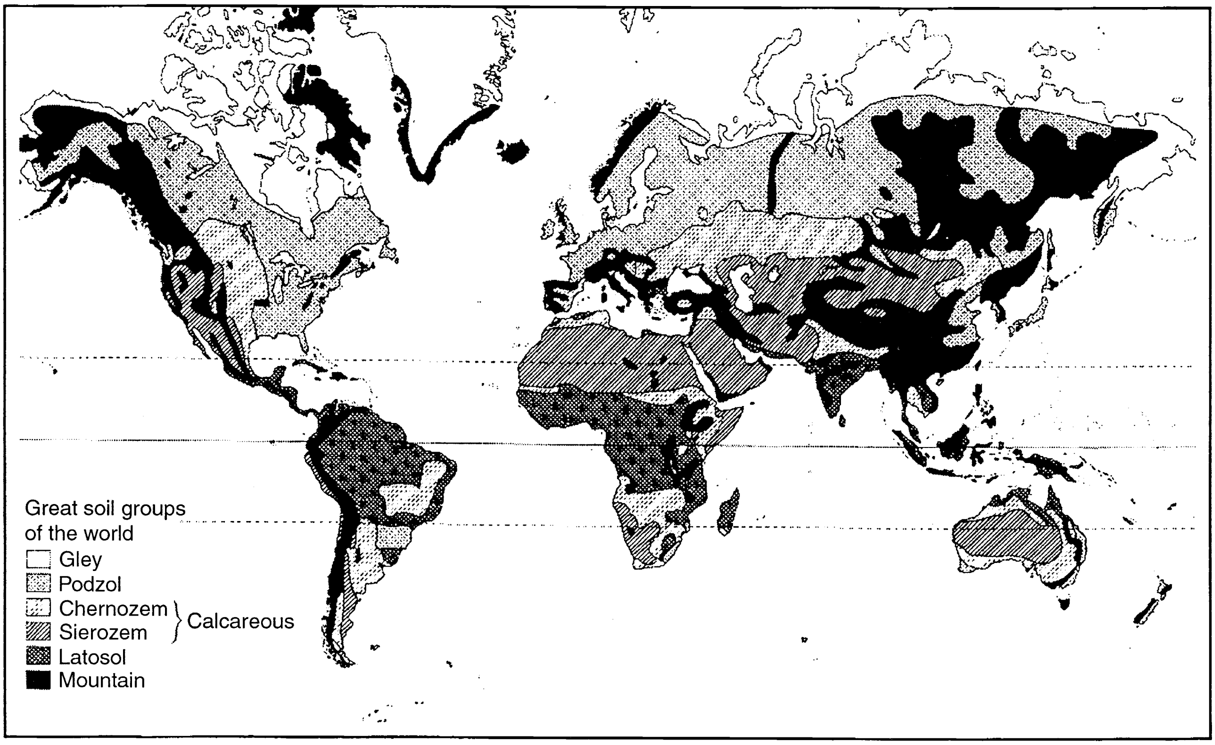


FIGURE 3 World distribution of major soil types. Note the close correlation of these soil types with the climatic zones shown in Fig. 2, reflecting the influence of temperature and precipitation on soil formation.

and most of them appreciated the fact that climatic conditions had changed, sometimes dramatically, during earlier periods of earth's history. Yet until the 1960s, most biogeographers clung to the belief that earth's landforms and ocean basins remained fixed. During the twentieth century, acceptance of the theory of continental drift and plate tectonics revolutionized the field of biogeography as much as acceptance of the theory of natural selection and evolution had in the previous century.

1. Continental Drift and Plate Tectonics

Although imperceptible to most of us, the earth's continents have moved, colliding at times and drifting apart at others: mountain ranges have formed and eroded away, seas have expanded and contracted, and islands have appeared and disappeared. These changes must have had profound effects on local and regional climates and, in turn, on the geographic distributions and variations of all forms of life on earth. As we will see in a subsequent section, the theory of continental drift and plate tectonics is a relatively recent advance. Yet, with the possible exception of the acceptance of the theory of natural selection, no other contribution has had more of an impact on the field of biogeography.

Plate tectonics is defined as the study of the origin, movement, and destruction of the earth's plates and how these processes have been involved in the evolution of the earth's crust. The theory of plate tectonics has achieved general acceptance among nearly all scientists and reigns as a unifying paradigm of both geology and biogeography. Yet until just three decades ago, relatively late in the development of these fields, champions of this theory were viewed as oddballs and heretics.

As with any other revolutionary theory in science, it is extremely difficult to pinpoint the origins of the theory of plate tectonics. The great geologist Charles Lyell entertained the idea during the 1830s and 1840s, but then abandoned it in favor of the accepted doctrine of the fixity of the continents and ocean basins. In their attempts to explain the affinities of biotas of now isolated continents, Lyell, Joseph Dalton Hooker,

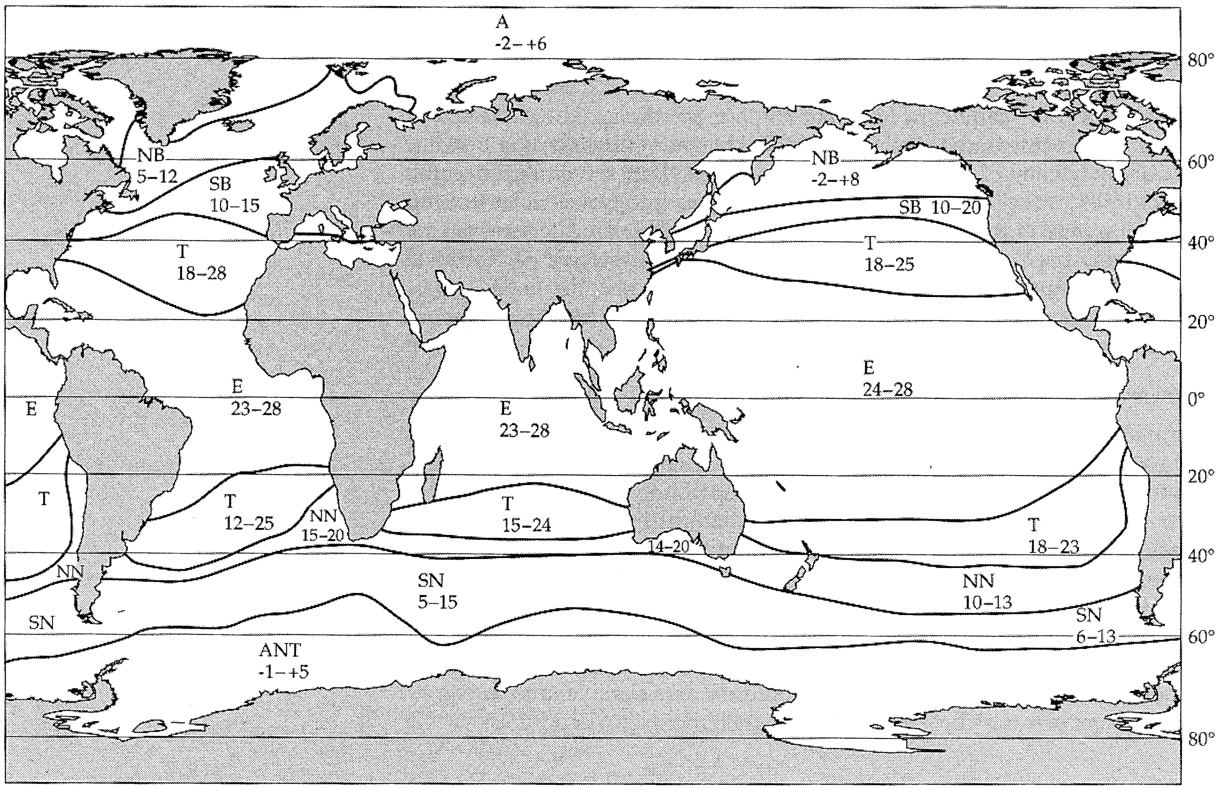


FIGURE 4 Climatic regions based on mean monthly water temperatures: A, arctic; NB, northern boreal; SB, southern boreal; T, tropical waters; E, equatorial region; NN, northern notal; SN, southern notal; ANT, antarctic. (After Rass, 1986.)

and other "extensionists" of the nineteenth century hypothesized the periodic emergence of great land bridges that then allowed biotic exchange. Darwin, Wallace, and other members of the dispersal camp soundly criticized such views: nothing vexed Darwin more than those extensionists who created land bridges "as easy as a cook does pancakes." In an uncharacteristically critical passage in one of his letters, Darwin complained to Charles Lyell of "the geological strides which many of your disciples are taking. . . . If you do not stop this, if there be a lower region of punishment for geologists, I believe, my great master, you will go there."

As it turns out, neither the extensionists nor the dispersalists were correct. In most cases, the similarities among now isolated biotas were instead the result of "dispersal" of the continents themselves. Perhaps the first important evidence for what was at first referred to as the theory of continental drift was the configuration of the continents. That is, once geographers had developed relatively accurate maps, it became clear to some that opposite coastlines seemed to fit. In 1858 one of Lyell's contemporaries, Antonio Snider-Pelligrini, may have been the first to demonstrate the geometric fit of the coastlines of continents on opposite sides of the Atlantic Ocean. Yet it wasn't until 1908 and 1910 that an American geologist, F. B. Taylor, and a German meteorologist, Alfred L. Wegener, independently developed models describing the movements of the earth's crust, along with the formation of mountain chains, island arcs, and related geologic features. Wegener continued to develop his model into a more comprehensive theory of continental drift, publishing his treatise in the 1920s. Wegener's theory, however, included too many assumptions about geologic processes and patterns that would not be well established for another three or four decades. His theory also included factual errors, such as overestimating the rate of movement of the earth's plates by perhaps two orders of magnitude. Finally, although he speculated on a potential mechanism, Wegener's theory really lacked a plausible one that could somehow drive the massive plates about the earth like bits of ice on a pond in spring. It is perhaps one of history's most tragic ironies that, in his quest to discover this mechanism by exploring a volcanically active region of Greenland, Wegener perished in a snow storm.

Wegener's insights would not be widely appreciated for another three decades. Acceptance of the theory of continental drift and its maturation to become the more comprehensive theory of plate tectonics would require many additional insights from geographers, paleontologists, and especially marine geologists during the 1940s and 1950s. These scientists found that when they "rejoined" the continents based on their geometric fit, not only did their biotas seem to match up, but so did topographic features such as mountain chains, rock strata, and fossil and glacial deposits. Perhaps most critical to the acceptance of the theory of continental drift were the efforts by marine geologists following World War II to map the surface of the ocean basins. It soon became clear that beneath each ocean lay a system of ridges that were situated far offshore. As one moved away from these ridges, the seafloor became deeper and more ancient as well. Provided with these and related clues, Herman Hess and his colleagues developed the theory of seafloor spreading: continental drift finally had an underlying mechanism (Fig. 5). Eventually, paleomagnetic evidence would allow marine geologists to estimate the previous positions of the continents and develop reconstructions of the sequences of their movements and creation and the dissolution of previous continents. Biogeographers were now armed with not just the evidence, but also the mechanisms responsible for the dynamics of biotas and the geographic template itself (i.e., immigration, extinction, evolution, and plate tectonics).

2. Glacial Cycles of the Pleistocene

The great shifting, collision, and separation of earth's plates profoundly affected the distribution of its biota, both directly and indirectly. Not only did plate tectonics alter major dispersal routes among biotas, but it substantially changed both global and regional climates. As plates shifted across different latitudes, their local biota was exposed to major shifts in climatic conditions. Areas of what is now tropical Africa, South America, and Australia once were situated over the south pole and exposed to severe antarctic climates.

On a global scale, drifting continents also triggered great shifts from periods of global warming to those dominated by glacial conditions. Land absorbs substantially more solar energy than does water. Thus, global climates tended to be warmer when landmasses were situated near the equator, but cooled as they shifted poleward.

Yet global climates can change substantially even during periods too short for substantial shifting of earth's plates. For example, during the Pleistocene (roughly the past 2 million years), earth experienced many climatic upheavals. Rather than being caused by any shifts in plates (which must have been minor given the relatively short period), these climatic shifts were caused by periodic changes in characteristics of the earth's orbit (referred to as Milankovitch cycles; Fig. 6). These changes significantly altered the total amount of solar energy intercepted by the earth, ultimately causing the climatic reversals of the Pleistocene. During full glacial periods, global temperatures dropped by as much as 6°C and most landmasses beyond 45° latitude were covered with glaciers often 2 to 3 km thick. Because so much water was tied up in the glaciers, sea levels dropped by 100 to 200 m, thus uniting long-isolated biotas via temporary land bridges. For example, during the last glacial maximum, the region of Southeast Asia and Malaysia was united with Sumatra, Java, and Borneo to form Greater Sunda, while Australia and New Guinea formed the island continent of Sahul (Fig. 7).

Winds, ocean currents, and precipitation patterns also changed substantially between interglacial and glacial periods. With each climatic upheaval, environmental regimes shifted across both latitudes and elevations. Regions such as the American Southwest, which is now dominated by desert and xeric grasslands, were once covered with coniferous forests. Warming and drying conditions that led to the current interglacial period dramatically reduced these forests and caused them to shift toward the mountain peaks, where cool and relatively humid conditions prevail.

These and other events must have profoundly influenced the distributions of most if not all biotas. As Brown and Lomolino (1998) summarize, however, all the complex biogeographic dynamics of the Pleistocene were triggered by three fundamental changes in the geographic template:

1. Changes in the location, extent, and configuration of principal habitats.
2. Changes in the nature of environmental regimes (combinations of temperature, seasonality, precipitation, and soil conditions).
3. The creation and dissolution of barriers associated with changes in sea level or elevational shifts in habitats.

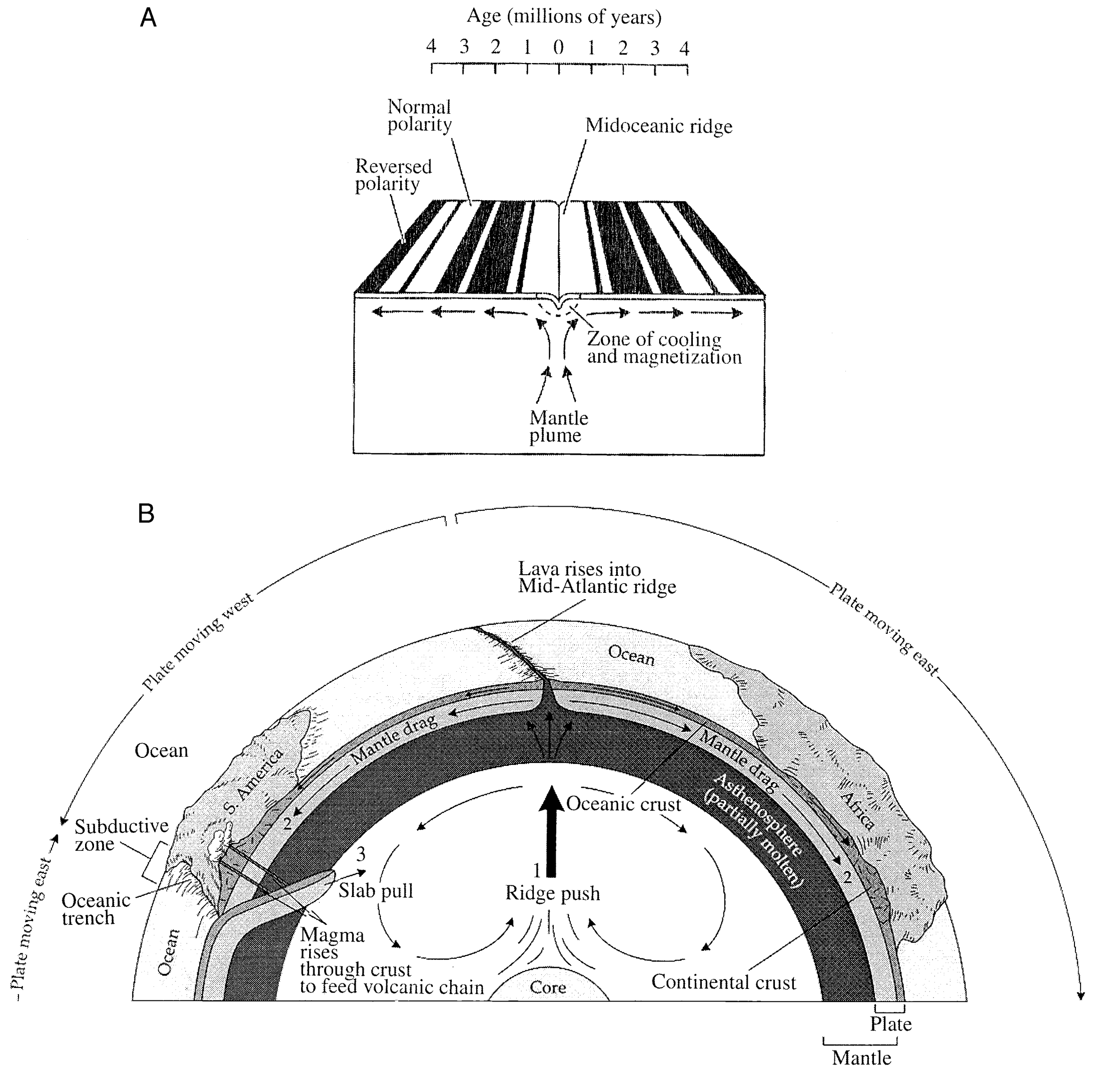


FIGURE 5 (A) During seafloor spreading, reversals in the earth's magnetic field are recorded as the magnetically sensitive, iron-rich crust cools. Differences in the widths of the magnetic stripes reveal differences in the duration of these polarity episodes and in the rate of seafloor spreading over time and among regions. (From Stanley, 1987.) (B) The current model of plate tectonics includes the possibility that at least three forces may be responsible for crustal movements: (1) ridge push, or the force generated by molten rock rising from the earth's core through the mantle at the midoceanic ridges; (2) mantle drag, the tendency of the crust to ride the mantle much like boxes on a conveyor belt; and (3) slab pull, the force generated as subducting crust tends to pull trailing crust after it along the surface. (After Stanley, 1987.)

The responses of both terrestrial and aquatic biotas, while no doubt complex, also were of three types:

1. Some species shifted geographically with their optimal habitats.
2. Some species remained and adapted to the altered local environment.
3. Other species, unable to modify their ranges or ecological associations, suffered range contraction and eventual extinction.

The biogeographic dynamics of the Pleistocene remains one of the field's most active and interesting study areas. Recent advances in analyzing and dating fossil material continue to add to our ability to reconstruct the historical development of biotas (Theme 2)

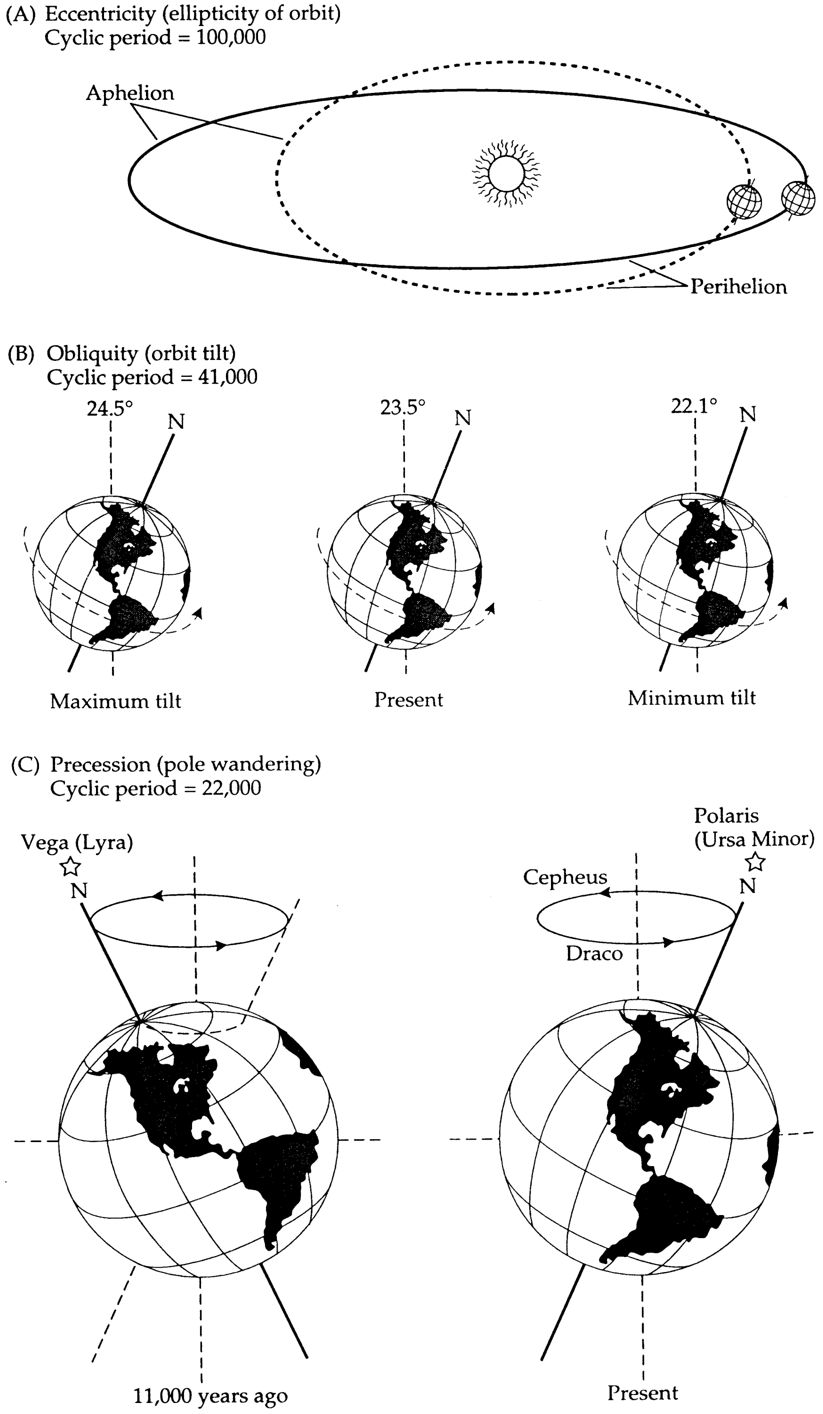


FIGURE 6 Milankovitch cycles are periodic changes in the eccentricity, obliquity, and precession of the earth's orbit. Each of these changes influences the earth's interception of solar radiation; therefore, these cycles may have been largely responsible for the glacial cycles of the Pleistocene. (After Gates, 1993.)

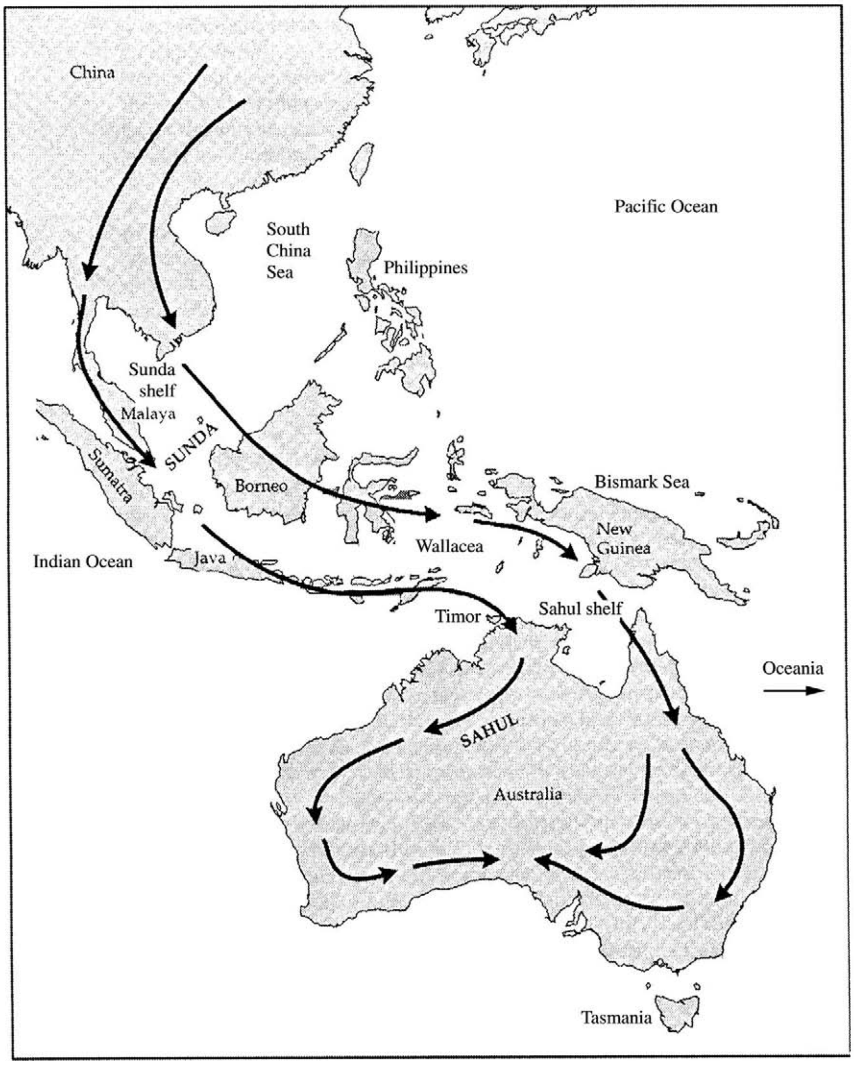


FIGURE 7 The lowering of sea levels during glacial maxima of the Pleistocene caused the exposure of continental shelves and the formation of dispersal routes across four regions of the eastern Pacific: Sunda, Wallacea, Sahul, and Oceania. (White areas = land exposed during glacial maxima; dark shading = deep water (>200 m); possible dispersal routes are indicated by arrows). (After Fagan, 1990; Guilaine, 1991.)

and, in turn, understand major episodes of biotic interchange and recent extinctions, as well as the current distributions of species.

B. Current Trends in Biogeography

1. Gradients in Species Diversity and Composition

In addition to biogeographic reconstructions, modern biogeographers continue to study an impressive diversity of patterns encompassing each of the four persistent themes of the field. During the middle of the twentieth century, many biogeographers focused on more general questions. Rather than dissecting and reconstructing the range of selected species, they examined trends in the total number of species, or what is often termed species richness.

Though many patterns in richness have been studied, two have received the lion's share of attention: the species–area and species–latitude relationships. Early explorers and naturalist of the seventeenth and eighteenth centuries noted the tendencies for species richness to increase with area of a region or island, and be higher for tropical versus temperate, subarctic, and arctic biotas. Armed with data from many hundreds of additional surveys and with a battery of sophisticated statistical tools, twentieth-century biogeographers confirmed the great generality of these patterns and developed some relatively simple models to explain those patterns. Often these models were equilibrial, assuming that species richness resulted from the combined but opposing effects of processes such as immigration into an area (which added species) and extinction (which decreased species richness). MacArthur and Wilson's equilibrium theory of island biogeography is perhaps the prototypic example of such a theory, and one that has dominated the field since its first articulation in the 1960s. Their theory was developed to explain both the species–area relationship and the species–isolation relationship (i.e., the tendency for species richness to decrease as one moves from near to more isolated islands). Simply stated, because immigration rates (the number of species new to an island) should decrease while extinction rate (loss of species already present) should increase as the island accumulates species, the island should eventually reach a level of richness at which immigrations balance extinctions. This equilibrial level of richness should vary among islands: decreasing with isolation because immigration rates are lower for more isolated islands, and increasing with island area because populations on larger islands should be less prone to extinction.

2. Biogeography in the Twenty-first Century

The equilibrium theory stimulated many studies in biogeography and related fields of ecology, and has served as the paradigm of island biogeography for some four decades. Yet an increasing number of biogeographers are beginning to question its utility as a modern paradigm. Species richness is often influenced by speciation and disturbances (e.g., major storms and tectonic events), processes not included in MacArthur and Wilson's original theory. Either the theory has to be expanded to include these processes, or it will be replaced by an alternative model—one that may eventually become the new paradigm of the field.

Whatever form such a model takes, it must be sophisticated enough to address the growing complexity of questions and patterns that we now study. MacArthur and Wilson's model was primarily developed to explain patterns in richness along gradients of area and isolation. Yet modern biogeographers are now searching for a theory to explain patterns across other geographic gradients and, perhaps more important, to explain geographic trends in the types rather than just the numbers of species. Why do the proportions of large versus small, endothermic versus ectothermic, herbivorous versus carnivorous animals, woody versus herbaceous, or annual versus perennial plant species vary across geographic gradients? How are biogeographic reconstructions related to phylogenies? In what manner do gene frequencies vary with isolation or across other geographic gradients? How do the size and shape of geographic ranges vary with latitude and among taxonomic groups, and how do population density and other demographic parameters vary across the range of a species?

We may have reached a point at which our questions and appreciation for the complexity of nature have become too sophisticated for the relatively simple models that have dominated the field since the 1960s. If this is true, biogeography may be on the verge of a major scientific revolution, one that may well rival those triggered by the seminal insights of scientists such as Charles Darwin, Alfred Russell Wallace, Alfred Wegener, Robert H. MacArthur, and E. O. Wilson.

III. BIOGEOGRAPHY AND THE CONSERVATION OF BIODIVERSITY

Biogeographers study both the patterns and processes influencing the geographic variation of nature. We study not just how many species occur in a particular area, but why more are there than somewhere else and which ones are likely to be shared among areas. We study and attempt to develop explanations for what are now termed "hotspots," regions of relatively high numbers and high endemicity of species. Biogeographers also study variation in the geographic template, including that associated with anthropogenic disturbances such as the spread of exotic species or the spatial patterns of deforestation. Many biogeographers study extinction and have demonstrated that it has a geographic signature: loss of species tends to be highest for the smallest and most isolated sites, namely, oceanic islands and fragments of once expansive habitats on the mainland. Given this, it becomes obvious that few disciplines could be any more relevant to understanding and preserving biological diversity than biogeography.

Our task, however, is far from a simple exercise of just applying what we already know. Indeed, only a small fraction (perhaps just 2 or 3%) of all extant species have been described, and we know precious little about the geographic distributions of most of those. What we do know often comes down to just general patterns for common species, but conservation biologists require detailed information on the rare species—those that may be the exceptions to most rules.

We have, however, made great progress in recent years in mapping and measuring the intensity (number of endemic species) of hotspots of biological diversity. In theory, these hotspots of biodiversity should receive the highest priority from conservation biologists, especially when they coincide with high levels of human activity. Yet even these approaches, generated by existing survey information and sophisticated geographic analyses, are based on a relatively limited number of surveys.

To develop more effective strategies for conserving global diversity, we still require a much more thorough understanding of the geographic variation of nature. A number of distinguished ecologists and biogeographers, including E. O. Wilson, have called for greatly accelerated efforts to map the diversity of life. With adequate support, within just a few decades we could greatly expand our knowledge of the distributions of most lifeforms and, eventually, contribute to their conservation as well.

See Also the Following Articles

DARWIN, CHARLES • DISPERSAL BIOGEOGRAPHY • HISTORICAL AWARENESS OF BIODIVERSITY • HOTSPOTS • ISLAND BIOGEOGRAPHY • SPECIES AREA RELATIONSHIPS • VICARIANCE BIOGEOGRAPHY

Bibliography

Briggs, J. C. (1987). *Biogeography and Plate Tectonics.* Elsevier, Amsterdam.

Brown, J. H. (1995). *Macroecology.* Chicago University Press, Chicago.

Brown, J. H., and Lomolino, M. V. (1998). *Biogeography.* Sinauer Associates, Sunderland, Massachusetts.

Carlquist, S. (1974). *Island Biology.* Columbia University Press, New York.

Darlington, P. J., Jr. (1957). *Zoogeography: The Geographic Distribution of Animals.* John Wiley & Sons, New York.

Fagan, B. M. (1990). *The journey from Eden: the peopling of our world.* Thames and Hudson, London.

Gates, D. M. (1993). *Climate change and its biological consequences.* Sinauer Association, Sunderland, Massachusetts.

Guilaine, J. (1991). *Prehistory: the world of early man.* Facts on File Publishers, New York.

MacArthur, R. H., and Wilson, E. O. (1963). *The Theory of Island Biogeography.* Monographs in Population Biology. Princeton University Press, Princeton, New Jersey.

Maurer, B. (1994). *Geographic Population Analysis: Tools for Analysis of Biodiversity.* Blackwell Scientific Publications, London.

Rass, T. S. (1986). Vicariance icthyogeography of the Atlantic ocean pelagial. In: *Pelagic Biogeography,* pp. 237–241. UNESCO Technical Papers in Marine Science, 49.

Rosenzweig, M. L. (1995). *Species Diversity in Time and Space.* Cambridge University Press, New York.

Sauer, J. D. (1988). *Plant Migration: The Dynamics of Geographic Patterning in Seed Plants.* University of California Press, Berkeley.

Sclater, P. L. (1858). On the general geographical distribution of the members of the class Aves. *Zoology* **2**, 130–145.

Simberloff, D. S. (1988). Contribution of population and community biology to conservation science. *Annu. Rev. Ecol. Systematics* **19**, 473–511.

Simpson, G. G. (1980). *Splendid Isolation: The Curious History of Mammals in South America.* Pergamon Press, Oxford, United Kingdom.

Stanley, S. M. (1987). *Extinction.* Scientific American Books, Inc., New York.

Strahler, A. N. (1973). *Introduction to Physical Geography,* 3rd ed. 468 pp. Wiley, New York.

Udvardy, M. D. F. (1969). *Dynamic Zoogeography.* Van Nostrand–Reinhold, New York.

Wallace, A. R. (1869). *The Malay Archipelago: The Land of the Orang-utan, and the Bird of Paradise.* Harper, New York.

Wallace, A. R. (1876). *The Geographic Distribution of Animals.* Macmillan, London.

Wallace, A. R. (1880). *Island Life, or the Phenomena and Causes of Insular Faunas and Floras.* Macmillan, London.

Wegener, A. (1966). *The Origin of the Continents and the Oceans* (translation of the 1929 edition by J. Biram). Dover Publications, New York.

BIOPROSPECTING

Nicolás Mateo, Werner Nader, and Giselle Tamayo
Instituto Nacional de Biodiversidad (INBio)

GLOSSARY

bionics Science of systems, which function in the same way as or similar to living systems.

biopiracy Illegal appropriation or exploitation of genetic and biochemical resources.

chemical prospecting Search for natural compounds in wild living plants, animals, and microorganisms with a potential for the development of chemical products like pharmaceuticals, pesticides, cosmetics, or food additives.

combinatorial chemistry Laboratory methods to produce all possible combinations from various sets of chemical building blocks in a short period of time and to generate molecular diversity for the screening of new bioactive compounds.

gene prospecting Search for genes in wild living plants, animals, and microorganisms for the breeding or genetic engineering of plants, animals, and microorganisms in agriculture, fermentation, and cell culture for agricultural and industrial production.

INBio The Instituto Nacional de Biodiversidad (National Biodiversity Institute), a Costa Rican association created in 1989 to generate and disseminate knowledge and promote sustainable uses of biodiversity.

BIOPROSPECTING IS THE SYSTEMATIC SEARCH for genes, natural compounds, designs, and whole organisms in wild life with a potential for product development by biological observation and biophysical, biochemical, and genetic methods, without disruption to nature.

I. BIOPROSPECTING: A TOOL FOR SURVIVAL AND A SOURCE OF INSPIRATION AND INNOVATION

A. Traditional Bioprospecting

In 1991 a 5300-year-old mummy was discovered in the Tyrolean Alps. The "Ice Man" and his effects were extremely well preserved and his clothing and equipment were perfectly functional for the survival under the harsh climatic conditions of the Alps. They consisted of plant and animal fibers, oils, and waxes, put together by a highly skilled and sophisticated Neolithic culture. The "Ice Man" suffered from an intestinal endoparasite, the whipworm *Trichuris trichiura*, but he was

Encyclopedia of Biodiversity, Volume 1

already equipped with the corresponding medicine, the fruiting body of *Piptoporus betulinus*, which contains oils that are toxic to parasites and compounds that act as strong laxatives that would cause expulsion of the dead and dying worms and their eggs.

The search for and utilization of biological resources is as old as humankind and has been key to the survival, adaptation, and evolution of the human species. The perceptions and values of communities and individuals depending on and living in close proximity to wild and domesticated biodiversity are clearly different from those of the modern scientist engaged in bioprospecting. To many communities, biodiversity entails magic, religious, and ceremonial connotations in addition to the more mundane food, shelter, and medicine benefits derived from its domestication and utilization.

The forest has provided nourishment and shelter to many cultures through a myriad of products and services since early times. The Neotropical forests harbor the wild ancestors of such major food crops as manioc, cacao, and yams. Evidence of squash (*Cucurbita moschata*) farming in southwestern Ecuador over 10,000 years ago suggests that agriculture in the Americas may have begun in the rain forests (Piperno and Pearsall, 1998). Low population pressure and markets limited to exchange and barter enabled the original indigenous groups to lead a low-impact lifestyle. Hunting and gathering, plus domestication of common beans, maize, cacao, chilies, squash, and other crops, permitted not only the survival of autochthonous groups but also their social and cultural development, and, quite significantly, with a minimal impact on the natural environment. The arrival of Europeans in the Americas brought along agricultural practices that were to a great extent destructive to Neotropical ecosystems. These new farmers felt compelled to “conquer” nature and cleared increasingly large areas of the forest to plant subsistence crops. Both systems have survived to the present, with the latter practices and large-scale, chemical-intensive agriculture being dominant because of increased population pressure and the need for higher agricultural productivity.

Costa Ricans, like other inhabitants of Middle America, have used and continue to use medicinal plants (e.g., roots of *Cephaelis ipecacuanha* and *Smilax* spp.), fruits and nuts, palms, birds, animals, felines (for hunting pleasure and to protect domestic animals), ornamental plants, feathers, and building materials from the forest. The impacts of these practices, particularly logging over the last 50 years, have been significant in altering the landscape and call into question the tropical region’s capacity for sustainable development.

In recent times, agronomists and professionals from related sciences have done considerable agricultural prospecting looking for landraces having agronomic advantages, particularly higher yields, pest and disease resistance, and adaptation to specific microhabitats. Gene prospecting of the resources managed, nurtured, and selected over centuries by indigenous communities has yielded considerable benefits to the world in terms of improved crop varieties and hybrids. Traditional plant breeding has slowed down in tropical countries due to lower research investment and unsuitable intellectual property protection, and this could diminish the opportunity of using the tools of the new biotechnologies.

Agricultural prospecting has resulted in the assembling of government, international, and private gene collections, however, the viability of these resources, mostly in tropical countries, has decreased over time. Some estimates suggest that over 70% of all germplasm collected since 1940 is no longer viable or the associated knowledge has been lost. *In situ* conservation, therefore, has become an attractive proposition to protect and utilize the remaining genetic resources.

Even plant breeding no longer depends only on genes from wild types and ancestors of modern crops; new improved varieties can be artificially generated through genetic engineering, for example, with designer genes that confer resistance to pests. Motifs in the amino acid sequences of the bactericidal and fungicidal magainin peptides from frog skin inspired genetic engineers to design resistance genes, which have scarcely anything in common with their natural ancestors (Nader and Hill, 1999).

B. Modern Bioprospecting

Modern science and technology have provided substitutes for some biological products: nylon, polyacryl, and polyester instead of cotton and wool; aniline dyes instead of natural colorants; and organophosphates or carbamates instead of plant-derived natural pesticides like rotenone, nicotine, and ryanodine. Yet human ingenuity still depends in great part on inspirations from wild life-forms to create these substitutes. In contrast to our forest-dwelling bioprospecting ancestors, we no longer need to use the biological original to create a final product; rather, through the study of these originals we are able to discover and unveil the principle of a biological function and then develop novel products on this basis.

Millions of years of evolution have created a wealth of structures and mechanisms at the molecular, cellular, and macrostructural level, all of which function economically and interact to perfection. Nature provides solutions to most of life's technical problems. Natural selection has imposed on living organisms the Min-Max Principle: a minimum of material and energy accomplishes a maximum of efficiency and stability. This makes biological prototypes particularly important for our future given the world's decreasing resources and increasing environmental problems (Hill, 1997).

Box 1

Biodiversity offers three fundamental sources of inspiration to the modern scientist: chemicals, genes, and designs. Fields of applications include drug development, agrochemistry, and cosmetics (chemicals), development of recombinant pharmaceutical proteins, enzymes, and agricultural biotechnology (genes), and architecture, mechanical engineering, and sensor technology (designs) (Nader and Hill, 1999).

An example of how nature inspires inventors is the development of the first glider airplane. By analyzing the flight of the stork (the biological prototype), the principle of airlift was discovered and the functional principle of flight was reduced to the vault of the wings. Vaulted wings derived from nature became the basic structural feature of all airplanes (Hill, 1997). Given the three inspirational resources from living nature—chemicals, genes, and designs—bioprospecting can be differentiated into chemical, gene, and bionic prospecting (Fig. 1). To discover new biological prototypes, bioprospectors can evaluate traditional uses of plants, animals, and microorganisms, use random sampling, or, in the biorational approach, systematically analyze biological phenomena (Tamayo *et al.*, 1997).

1. Chemical Prospecting

The major defense mechanisms of plants against herbivores rely on chemicals: "The world is not colored green to the herbivore's eyes, but rather is painted morphine, L-DOPA, calcium oxalate, cannabinol, caffeine, mustard oil, strychnine, rotenone, etc." (Janzen, 1975). The coevolution of herbivores and their feeding plants has created diversity on both the species and the molecular levels (Ehrlich and Raven, 1964). Communication, competition, sexual attraction or rejection, and pollination are also based to a great degree on chemistry and have contributed to the development of diversity. To help conserve this chemical wealth for future generations, Thomas Eisner developed the concept of chemical prospecting, a collaborative effort among conservationists, scientists, the pharmaceutical industry, and biodi-

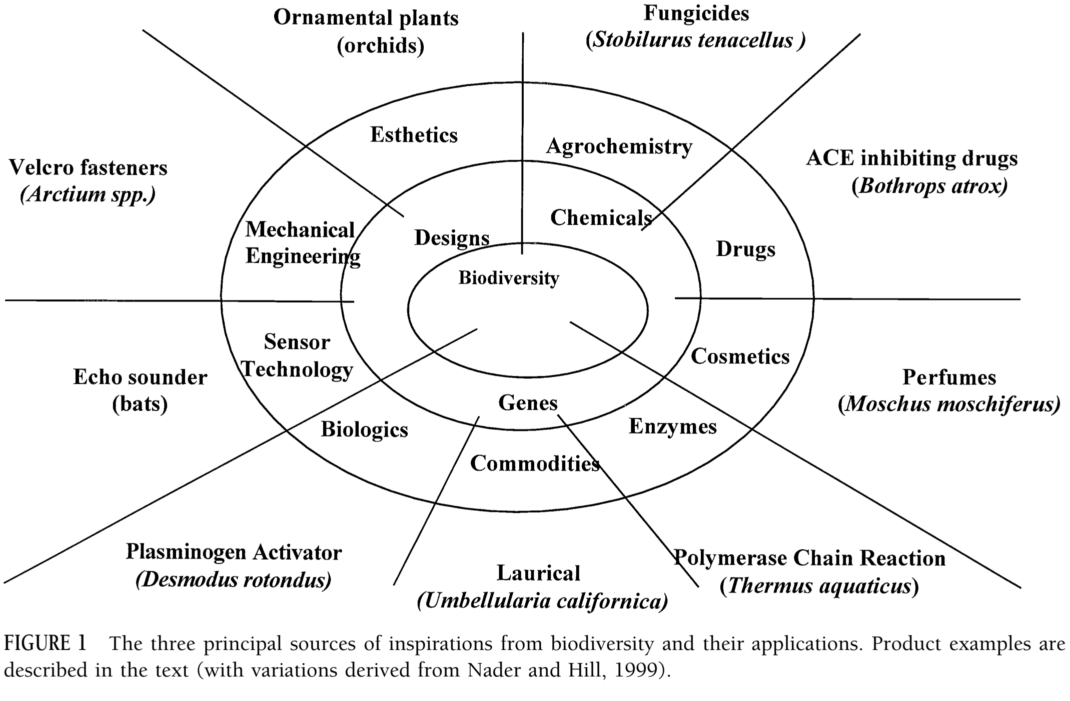


FIGURE 1 The three principal sources of inspirations from biodiversity and their applications. Product examples are described in the text (with variations derived from Nader and Hill, 1999).

versity-rich countries to develop products from biodiversity and to generate income for its conservation (Reid *et al.*, 1993).

Although human ingenuity has created a huge variety of chemical compounds with exceptional pharmaceutical activity without inspiration from nature—like the anxiolytic benzodiazepines valium and librium—natural compounds and their derivatives are still of central importance in drug discovery. Table I lists the 30 best-selling pharmaceuticals (drugs and biologics) in 1997. Thirteen were either derived from or developed as the result of leads generated by natural products from wild plants, animals, and microorganisms. The antihypertensive proline derivatives enalapril (Vasotec) and lisinopril (Zestril) were derived from a peptide in the venom of the fer-de-lance (*Bothrops jararaca* or *B. athrox*). Diclofenac (Voltaren) belongs, like aspirin, to the group of nonsteroidal anti-inflammatory drugs, which were derived from the lead structure salicin from willow bark (*Salix* spp.). The antineoplastic paclitaxel (Taxol) was discovered in the screening program of the National Cancer Institute from the bark of the Pacific yew tree (*Taxus brevifolia*). The antiviral drug acyclovir (Zovirax) was developed using prior knowledge of cyto-

TABLE I

Ranking of Key Ethical Pharmaceuticals (Drugs and Biologics) by Sales Value (U.S. $) in 1997[a]

Rank	Product	Product category	Company	U.S. $ (millions)
1	Losec/Prilosec	Proton pump inhibitor	Astra/Astra Merck	3749
2	**Zocor**	**Hypolipidemic**	**Merck & Co.**	**3575**
3	Prozac	Antidepressant	Eli Lilly	2559
4	**Vasotec**	**ACE inhibitor**	**Merck & Co.**	**2510**
5	Zantac	H2-antagonist	Glaxo Wellcome	2254
6	Norvasc	Calcium antagonist	Pfizer	2217
7	Claritin	Antihistamine	Schering-Plough	1726
8	**Augmentin**	**Penicillin antibiotic**	**SmithKline Beecham**	**1516**
9	Zoloft	Antidepressant	Pfizer	1507
10	Seroxat/Paxil	Antidepressant	SmithKline Beecham	1474
11	**Pravachol**	**Hypolipidemic**	**Bristol-Myers Squibb**	**1437**
12	Ciprobay	Quinolone antibiotic	Bayer	1378
13	*Premarin*	*Hormonal product*	*American Home Products*	*1328*
14	**Mevalotin**	**Hypolipidemic**	**Sankyo**	**1311**
15	**Biaxin/Klaricid**	**Macrolide antibiotic**	**Abbott**	**1300**
16	*Novolin*	*Human insulin*	*Novo Nordisk*	*1290*
17	Pepcid	H2-antagonist	Merck & Co.	1180
18	**Sandimmun/Neoral**	**Immunosuppressant**	**Novartis**	**1166**
19	*Epogen*	*Antianemic*	*Amgen*	*1161*
20	**Mevacor**	**Hypolipidemic**	**Merck & Co.**	**1100**
21	Imigran	Antimigraine	Glaxo Wellcome	1085
22	*Procrit/Eprex*	*Antianemic*	*Johnson & Johnson*	*1070*
23	*Neupogen*	*Colony stimulating factor*	*Amgen*	*1068*
24	Adalat	Calcium antagonist	Bayer	1063
25	**Rocephin**	**Cephalosporin**	**Roche**	**1062**
26	Prepulsid	Gastroprokinetic	Johnson & Johnson	1050
27	**Zestril**	**ACE inhibitor**	**Zeneca**	**1036**
28	**Voltaren**	**NSAID**	**Novartis**	**1019**
29	**Zovirax**	**Antiviral**	**Glaxo Wellcome**	**951**
30	**Taxol**	**Anticancer**	**Bristol-Myers Squibb**	**941**

[a] Source: Wood Mackenzie's PharmaQuant, Edinburgh, U.K., Jan. 1999. Products derived from natural compounds from wildlife animals, plants, and microorganisms are printed in bold, and products derived from natural compounds are in italic.

sine arabinoside, which was isolated from a Florida sponge (Alan Harvey, pers. comm.). Lovastatin (Mevacor) and its synthetic analog simvastatin (Zocor) are derived from the fungus *Aspergillus terrestris*, and antihypercholesterolemic drugs like pravastatin (Pravachol or Mevalotin) are natural compounds from *Penicillium citrinum*. Ceftriaxone (Rocephin) is a semisynthetic third-generation cephalosporin antibiotic (*Cephalosporium* sp.) and cyclosporin (Sandimmun/Neoral) is an immunosupressant cyclic peptide from *Tolypocladium inflatum*. Co-amoxiclav (Augmentin) is a combination of the beta-lactamase inhibitor clavulanic acid from the bacterium *Streptomyces lavuligerus* and the semisynthetic antibiotic amoxicillin (*Penicillium* or *Aspergillus* spp.). Finally, clarythromycin (Biaxin/Klaricid) is a chemical derivative of the classic antibiotic erythromycin from the streptomycete *Saccharopolyspora erythrea*.

Random screening for new drugs from natural compounds comprises the collection of materials from plants and animals and the isolation of bacteria and fungi in large numbers. Extracts are prepared from these materials and then exposed to screening batteries with automated bioassay systems to discover novel biological activities. The bioactive principles are then purified and characterized by bioassay-guided fractionation. Since a large number of bioactive compounds will be already known and thus lack novelty, these need to be identified as early as possible in the discovery process by dereplication. Even if a compound is in the public domain, it might still be patentable in connection with a novel application. Another drug discovery route is provided by chemical screening, which fits the demand of modern high-throughput screening for large quantities of purified compounds. Without knowing their biological activity, extracts are first fractionated into pure compounds and these are then exposed to the bioassays.

Among the most notable developments in modern medicine were acetylcholinesterae (ACE)-inhibiting drugs like enalapril (Fig. 2a), captopril (Fig. 2b), lisinopril, and perindopril. These compounds were derived biorationally from a peptide in the venom of a deadly tropical snake (Wyvratt, 1988). The fer-de-lance (*Bothrops athrox* or *B. jararaca*) is a Neotropical pit viper that kills its prey by causing a drop in blood pressure (in the context of pharmaceutical development, this represents the observation of a biological prototype). Hypertension is a major health problem in most industrialized countries (the formation of an analogy between a medical problem and a biological phenomenon). A peptide in the venom inhibits an enzyme in the mammalian bloodstream, which converts the peptide angiotensin I to the hypertensive hormone angiotensin II. Analysis of the functional principle of this snake peptide at the molecular level and the discovery of the bioactive lead structure (a proline in combination with the carboxyl group of a glutamic acid) finally led to the synthesis of small molecules like captopril and enalapril (the technical realization of a biological principle). Thanks to drugs like the ACE inhibitors and, especially, these deadly snakes from the Neotropics, human deaths from stroke and heart attack have decreased by over 50% during the last two decades.

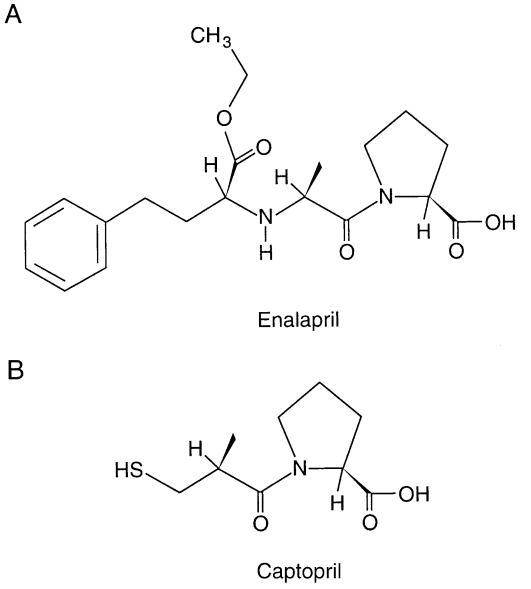


FIGURE 2 (A) The antihypertensive drug enalapril (Merck & Co.). (B) The antihypertensive drug captopril (Bristol Myers Squibb).

Many of today's pharmaceutical industry multinational corporations became industrial giants by producing and selling products derived from ethnobotanical research. Prior to Bayer's aspirin, American and Eurasian peoples treated fevers, inflammation, and pain with salicin-containing plants like willows and poplars. Novartis' diclofenac, number 28 in the list of top-selling pharmaceuticals of 1997 (see Table I), was derived from the same salicin lead. Other examples are stabilizing muscle relaxants like tubocurarine, alcuronim, and pancuronium, which are routinely used during anesthesia. They were derived from curare, an arrow poison prepared from toxic plants by indigenous people in the Amazon and Orinoco valleys of South America. The anticancer drugs etoposide and teniposide are derivatives of podophyllotoxin, a glycoside isolated from the roots of various species of the genus *Podophyllum*. These plants possess a long history of medicinal use by early American and Asian cultures, including the treatment of skin cancers and warts. Cholesterol-lowering drugs like lovastatin, simvastatin, and pravastatin now have medical competition from a yeast extract from fermented Chinese red rice. Asians use these extracts traditionally as a food additive and the yeast produces lovastatin naturally.

Secondary metabolites from plants have been used traditionally in agriculture as natural pesticides, but they were pushed out of the market to a great extent following the advent of DDT. Other natural pesticides are nicotine from tobacco leaves, rotenone from roots of the genus *Derris*, quassin from the wood of *Quassia amara*, ryanodine from leaves, stems, and roots of the tree *Ryania speciosa*, azadirachtins from the Asian neem tree, and pyrethrins from the perennial and *Chrysanthemum cinerariifolium*, originally from Dalmatia and Montenegro (Wink, 1993). However, the toxicity of chemically synthesized pesticides like organophosphates and chlorinated bicyclic compounds soon became evident and derivatives of natural insecticides, fungicides, and herbicides have gained increasing importance since then. Examples are the pyrethroids, the fungicides kresoxim-methyl (Fig. 3b) and azoxystrobin (Fig. 3c), and the herbicide phosphinothricin. Pyrethroids are derivatives of the pyrethrins, which were used traditionally as insecticides. Kresoxim-methyl and azoxystrobin are derived from strobilurin A (Fig. 3a), a natural fungicide that was discovered via biorational research on the mushroom *Stobilurus tenacellus*. Random screening led to the discovery of phosphinothricin in the culture supernatants of a streptomycete (Crouse, 1998).

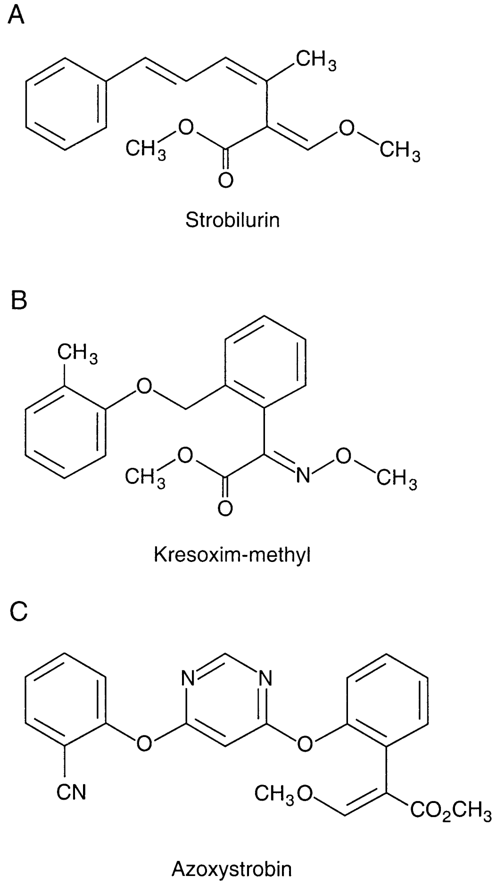


FIGURE 3 (A) The natural fungicide strobilurin A from the mushroom *Stobilurus tenacellus*. (B) The strobilurin class fungicide kresoxim-methyl (BASF). (C) The strobilurin class fungicide azoxystrobin (Zeneca).

2. Gene Prospecting

The selection of various phenotypes (and thus genes) has been used traditionally for plant and animal breeding, and modern gene technology offers many new opportunities. Numerous products derived from genes from wild life are already on or are close to entering the market (Tamayo *et al.*, 1997).

A powerful protein to treat acute heart attack was discovered biorationally in the saliva of the common vampire bat, *Desmodus rotundus*. While feeding on the blood of their victims, these bats release analgesics and coagulation inhibitors with their saliva, including the Desmodus Plasminogen Activator (DPA), which dissolves thrombolytic blood clots and allows clot-free drinking. Recombinant human tissue Plasminogen Activator (tPA) has been approved as a therapeutic agent against heart attack in the United States and Europe, although researchers at Schering AG found that DPA is more efficient and safer for therapeutic application. Consequently other blood-sucking organisms like ticks, mosquitoes, and hookworms are currently under investigation in attempts to discover novel anticlotting, antiplatelet, and vessel-constricting principles.

One example of the value of traditional knowledge is the development of the recombinant pharmaceutical protein hirudin, a potent anticoagulant derived from the saliva of leeches (*Hirudo medicinalis*). Leech therapy has been used in traditional medicine in Europe and Asia for centuries as the only efficient treatment of thrombosis, thrombophlebitis, and hypertension.

Traditionally, humans have used enzymes for the production of cheese, softening of leather, and fermentation of tea and tobacco. In modern technology, enzymes have gained increasing importance in chemical engineering, in the processing of animal and human food and textiles, as a component of laundry detergents, and for the diagnosis of diseases. Gene prospecting searches for enzymes with new substrate specificity and physicochemical characteristics, and gene technology allows their production in nearly unlimited amounts at low cost. Today random screening can be applied in the discovery process at high efficiency rates and without the need to isolate and cultivate microorganisms. DNA is isolated directly from water, soil, dung, carcass, or compost, then randomly cloned into expression libraries; the host bacteria are subsequently screened for enzyme activity in high-through-put screening robots on microtiter plates. From the hosts, the cloned DNA can be directly amplified and sequenced, and the genetic information used to construct highly productive host strains for the bioreactors.

Hydantoinases generate D-amino acids, high-value fine chemicals for the production of half-synthetic antibiotics and pesticides. Particularly useful are thermostable hydantoinases, which were discovered by random screening of cultivated thermophilic bacteria from hot springs in Yellowstone National Park in the United States. Yellowstone was also the source of Taq polymerase from *Thermus aquaticus*, a crucial component of the polymerase chain reaction (PCR). Its inventor, Kary B. Mullis, plucked *T. aquaticus* from the American Type Culture Collection, where the microbiologist Thomas D. Brock from the University of Wisconsin had deposited it in the late 1960s. Recently some users of PCR have replaced Taq polymerase by Pfu polymerase, which was isolated from *Pyrococcus furiosis* ("flaming fireball") and works best at 100°C. Proteases and lipases from psychrophilic bacteria, which decay whale cadavers in the deep sea, are of interest for producers of cold-wash laundry detergents (Madigan and Marrs, 1997).

Food for poultry and pigs is increasingly processed with phytases from soil fungi like *Aspergillus niger*. This enables these monogastric animals to assimilate the phosphorous from plants and diminishes the amount of phosphorous that enters the environment in their manure. In the United States alone, the widespread application of phytase could prevent 82,000 tons of phosphate from entering the environment each year and contributing to the eutrophication of rivers and lakes.

Genes from wild plants, vertebrates, insects, and soil bacteria are used to create transgenic crops that carry resistance to pests and herbicides. They can be discovered biorationally, by random screening, or on the basis of knowledge from traditional uses.

Although frogs live in ponds infested with bacteria and fungi, they rarely get infected by these pathogens. This common knowledge led to the discovery of antimicrobial peptides, the magainins, which can be genetically expressed in transgenic plants and which confer resistance to fungi and bacteria. Random screening of actinomycetes from the soil of Cameroon led to the discovery of phosphinothricin acetylase, which confers resistance to the herbicide Basta in transgenic crops like soy bean, rape, and cotton. Resistance to insects can be achieved by the expression of insecticidal proteins from soil bacteria, like delta-endotoxin from *Bacillus thuringiensis* and cholesterol oxidase from *Streptomyces* sp. (Tamayo *et al.*, 1997).

As a consequence of their absorptive mode of nutrition, fungi are able to exploit an almost infinite diversity of nutritional microniches. Recent studies suggest that considerable fungal diversity exists even in what would be considered a homogenes substrate. For example, Bills and Polishook (1994a) adapted the particle filtration technique to maximize the diversity of fungi isolated from a substrate. Whereas traditional isolation techniques select for fast-growing species that utilize primarily simple carbohydrates, the particle filtration technique favors isolation of those fungi actively growing in the substrate. In their study, fungi were isolated from 1-ml particle suspensions of tropical rain forest leaf litter. From each sample, 80–145 different fungal species were isolated and the four replicates of the same substrate yielded between 300 and 400 different species of fungi. In a second study (Bills and Polishook, 1994b), fungi were isolated from only one vascular plant host, *Heliconia mariae* J. D. Hooker. From 0.8-ml suspensions of filtered particles of leaves from four plants, they isolated about 50–100 fungal species per leaf sample. In both of these studies, approximately 40–60% of the fungi isolated were rare or unidentifiable. Clearly the fungi are far more diverse than is currently recognized, even by mycologists.

Metabolic pathway engineering also offers new agricultural opportunities. A gene for a 12-carbon thioesterase that transforms the common rape plant into a producer of high-laurate canola oil has been discovered in the undomesticated California bay (*Umbellularia californica*). This oil is a perfect substitute for palm kernel and coconut oil, which are high-value commodities from tropical countries. Already 70,000 acres of transgenic rape are under cultivation in the United States and Canada.

Developments like Laurical (the trade name for this high-laurate canola oil) displace the production of tropical commodities to the farm fields of developed countries, or even possibly to industrial bioreactors, posing a threat to the economic survival of millions of farm families in less developed countries. In Africa alone, $10 billion in annual exports are vulnerable to industry-induced changes in raw material prices and requirements. Most developments in plant biotechnology have been achieved with crops mainly cultivated in industrialized countries. This fact reduces the likelihood that farmers in developing countries will benefit from the new agricultural opportunities of gene technology (Tamayo *et al.*, 1997).

3. Bionic Prospecting

Architects, designers, and engineers use prototypes from nature for construction and technical solutions. Josef Baxton, after seeing the radial ribbing in the venation of the Amazonian water lily (*Victoria amazonica*), designed on this basis the dome of the crystal palace in London. The inventor of Velcro fasteners was inspired by the stickiness of the seeds of a common farm weed, the burdock (*Arctium* spp.). The development of the echo-sounder was inspired by the ultrasonic orientation system of bats, and infrared sensors were based on the design of the thermosensitive pit organ of the rattlesnake. The latter developments are of particular importance for military use and thus it is not surprising that the concept of "bionics" was created in 1958 by the Air Force major Jack E. Steele as the "science of systems" that function in the same way or similar to living systems (Hill, 1997).

The discovery of new technical principles in living systems is far from being completed, and at least one major resource is still untapped: the insect world of the tropics. Studies by Hill *et al.* (in press) in Costa Rica on butterflies and grasshoppers have revealed a diversity of new coupling mechanisms and led to the definition of "tropical bionics" as a new branch in engineering. Insects became the most species-rich group of eukaryotes by their capability to adapt to ecological niches. They are not only excellent chemists and genetic engineers, but also masters in mechanical engineering, lightweight construction, and sensor technology. Figure 4 shows some of the morphological and physiological characteristics of a butterfly. Because of their diversity, insects are an enormous reservoir of potential technological developments.

Until recently, the development of technical solutions in engineering on the basis of biological prototypes was mainly driven by biorational observations. For example, sharks have a rough skin, and common sense suggests that rough surfaces should increase the resistance in water and thus slow the fish down. However, sharks swim at speeds of over 50 km/hr. This paradox led to the discovery that the structure of the shark squamae impedes microturbulence, which commonly emerges on smooth surfaces at high speed. The technical development of this principle led to the creation of new surfaces for airplanes that significantly reduce kerosene consumption. Another example of the biorational approach comes from the lotus, considered to be the flower of purity and cleanliness in Asian cultures. Tiny wax papillae on its surface act as a self-cleaning mechanism. This principle is currently being

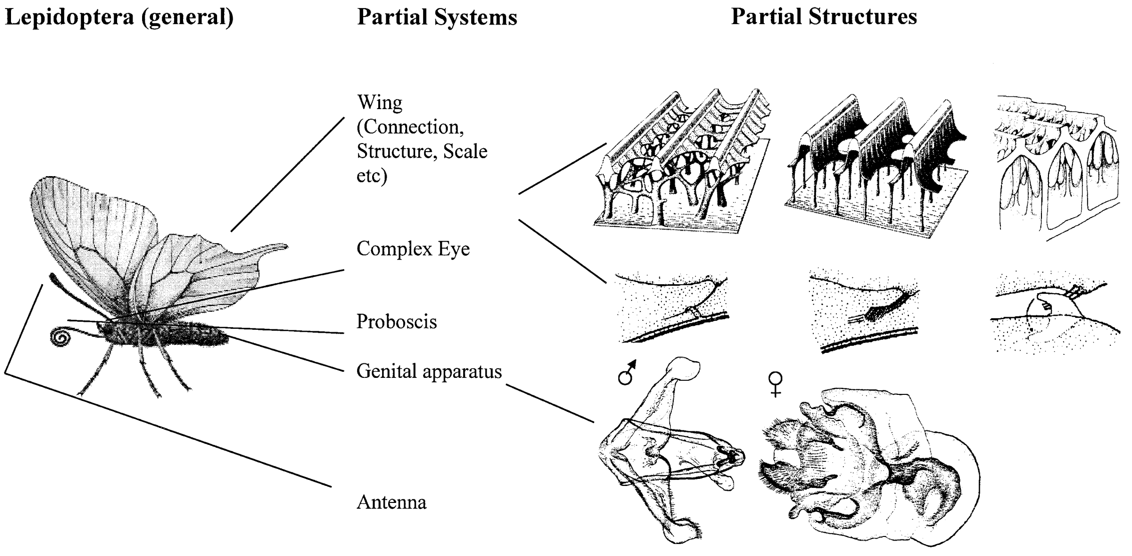


FIGURE 4 The butterfly: a bionic system and technical prototype (with variations derived from Hill *et al.*, in press).

used to develop self-cleaning surfaces for buildings and cars.

A new tool for the discovery and development process in bionics was recently developed by Hill (1997). Structures and mechanisms discovered in nature are systematically listed in catalogs and organized after the principal functions in engineering: connection, separation, formation, carriage, support, transmission, storage and blocking of materials, energy, and information. These catalogs of inspirations are available to the engineer to screen for technical solutions, the blueprints of which nature has already provided.

II. ECONOMIC VALUE AND BENEFIT SHARING

A. Biodiversity-Related Markets

Markets related to products derived from biodiversity are significant, as shown in Table II. This table does not include bionic applications like sensor technology and robotics.

The estimated growth of the world drug market is about 6% per year, from 1997 sales of U.S. $295 billion to $378 billion in the year 2001 (Thayer, 1998). To become a billion-dollar product, a newly developed pharmaceutical has to be better than similar products already on the market or reveal a new mode of action. In the highly competitive drug market, innovation and production speed are paramount, and consequently in the United States R&D costs increased within ten years from U.S. $6 billion to $20.6 billion in 1998. The supply of screening batteries for the synthesis and extraction of new compounds consumes 11% of these R&D budgets and requires a workforce of 4400 professionals or 12.9% of all employees in R&D (Thayer, 1998). To compete with synthetic and combinatorial chemistry efforts, bioprospecting needs to provide large numbers of natural compounds at high speed, for example, by chemical screening, or untap new and unusual resources.

TABLE II

Markets of Various Products and Services

Market	U.S. $ (billions)
Drug market (1997)[a]	295
Natural product-based drugs (1993)[b]	50
Phytomedicines (1993)[c]	12.4
Pesticides[d]	30
Seeds (estimate for 2000)[d]	50
Horticulture (U.K., 1991)[b]	1.6
Enzymes (1996)[e]	2.5
Cosmetics (U.S.A., 1994)[b]	20
Natural cosmetics[b]	0.5
Tourism (1995)[b]	3400
Nature tourism (1988)[b]	12

[a] From Thayer (1998).
[b] From ten Kate (1995).
[c] From Grünwald (1995).
[d] From K. ten Kate, pers. comm.
[e] From Madigan and Marrs (1997).

Global sales of U.S. $30 billion in 1995 makes agrochemistry the second largest biodiversity-related market. Because of the toxicity and environmental problems associated with synthetic chemicals, pesticides derived from natural compounds currently dominate the market (see earlier discussion), however, development of resistance in pests continually requires new and innovative products and creates demand for bioprospecting. Crop resistance to pests can be achieved by genetic engineering and this might, at least partially, substitute for some pesticide use in the future. The market for the seeds of these genetically engineered varieties is estimated to grow to U.S. $50 billion in the year 2000. Currently most genetic transformation is done with genes coding for insecticidal proteins from *Bacillus thuringiensis*. However, it is only a matter of time before crop resistance is overcome by pests, and therefore the discovery of new resistance genes from wild sources has become a central focus of bioprospecting efforts.

Other markets related to biodiversity are gaining increasing importance, like those for enzymes (see earlier), phytopharmaceuticals, seeds, and natural cosmetics. The worldwide market for phytomedicines derived from traditional medicinal sources is estimated at U.S. $12.4 billion, headed by products derived from ginseng, ginkgo, garlic, horse chestnut, and echinacea (Grünwald, 1995). Nutraceuticals and functional foods offer a health advantage in addition to nutrition. Examples are Chinese red rice and cholesterol-lowering oils, the latter enriched in plant sterols and stanol, both by-products of the paper pulping industry and corn wet-milling operations. With an estimated market of $7.9 billion in 1997, these products are expected to compete effectively with cholesterol-lowering drugs like lovastatin. In general, the future market looks bright for products derived from nature, with an increasing demand for natural compounds and wild genes for use in genetic engineering. The same is true for bionics.

Recently, attempts have been made to put price tags on natural resources, including biodiversity (Constanza *et al.*, 1997). The value of sources of unique biological materials and products was estimated as U.S. $79 per hectare of forest per year. The value of yet undiscovered pharmaceuticals in tropical forests was estimated at U.S. $3–4 billion for a private pharmaceutical company, and as much as U.S. $147 billion dollars to global society as a whole. Although significant, these numbers do not reflect the real value of inspiration that can be derived from wild plants, animals, and microbes. Biodiversity provides the goods and services that are essential in supporting every type of human endeavor and thus enables societies to adapt to different needs and situations. Biological prototypes may be crucial for the survival of mankind in the future, and the "Great Hunger" in Ireland in the mid-1800s taught a bitter lesson. This potato late blight (*Phytophthora infestans*) caused the death of 1 million and emigration of another 1.5 million people.

B. Bioprospecting and Conservation

The initial meaning of prospecting was the search for precious metals and oil in the soil and subsoil. Then Thomas Eisner, from Cornell University, coined the term "chemical prospecting," and more recently bioprospecting has been used to indicate the search for new sources of chemical compounds, genes, proteins, microorganisms, and other products that have economic potential and that can be found in the world's diversity of plants, animals, and microorganisms.

For bioprospecting to be an ally of conservation, it must be based on sound principles and criteria, and must be linked to the concept of value. It was highlighted earlier that the present and potential value of bioprospectible designs, genes, and chemicals is enormous and justifies the use of intelligent conservation strategies. To be effective as an element of conservation, bioprospecting must contribute to the generation of wealth through R&D and must also alleviate poverty in biodiversity-rich regions. In this context, bioprospecting programs at the national level can help to develop national capacity and provide economic returns to conservation projects if investments are made in science, technology, and market research. However, this may not be sufficient to achieve sustainable development: communities inside or bordering protected areas must participate in conservation and must benefit from bioprospecting activities in their own regions. The utilization of biodiversity linked to added value at the local level appears to be a promising avenenue to generate economic growth at the community level and to create incentives for conservation.

Economic benefits at the community level could include: growing, grading, and packing of medicinal plants; cultivation of new ornamental plants; domestication, cultivation, and primary production of biopesticides; cultivation and processing of aromatic teas; and exploration and evaluation of local landraces with resistant genes of interest. In all cases, a critical prerequisite for engaging in added-value biodiversity development at the community level is market research. Unless a clear market is identified, proper contracts are signed, and organization at the local level is achieved, there is no sense in proceeding with bioprospecting activities.

C. Biopiracy

Our world would be a totally different place today if crops and domestic animals had not been moved from their centers of origins and domestication to other parts of the world. Coffee to the Americas, potatoes to Europe, corn to Africa, and wheat to China are just a few examples of how germplasm has become the "patrimony of humanity." The ratification of the Convention on Biological Diversity (CBD) and the development and implementation of property rights regimes have changed completely the prior existing scenarios on the spread and use of germ plasm.

Today, in many cases genetic and biochemical resources are jealously guarded and are the cause of conflict (when appropriate and legally sound negotiations are absent) in many parts of the world. The term biopiracy has been coined to reflect the illegal appropriation or exploitation of genetic and biochemical resources. It must be added, however, that the lack of proper codes of conduct, lack of legislation, and the lack of national capacity to handle biodiversity utilization issues in the majority of biodiversity-rich countries are factors encouraging biopiracy. Even national researchers, when dealing with poorly defined or cumbersome legal procedures, tend to take shortcuts in their quest for useful genetic and biochemical resources.

D. Benefit Sharing Models and Bioprospecting Experiences

Perhaps the most controversial issue related to bioprospecting in tropical developing countries involves benefit sharing. This is true for the case of the Instituto Nacional de Biodiversidad (INBio) in Costa Rica, as well as for other models and experiences developed elsewhere.

The central questions are: What is equitable? Who benefits and how? Who makes the "rules"?

The issue of equitability is controversial and first arose with the initial perception that millions of dollars would flow back to the countries, organizations, and individuals involved in bioprospecting. The experiences of the last few years indicate that monetary benefits (unless royalties would materialize) to host countries, although significant, are limited in comparison to other less tangible benefits such as technology transfer, increased local scientific expertise, improvements in legal frameworks, and enhanced negotiating capacities. These less tangible benefits may be poorly understood or underappreciated by some segments of society, who quite rightly are interested in achieving a direct flow of resources and economic benefits to the local communities living near conservation areas.

Conditions are different in every country and therefore local solutions should be tailored to local circumstances. Inside Costa Rica's governmement-protected wildlands, for example, there are no human settlements, in contrast to the case in many other biodiversity-rich countries, and therefore straightforward agreements developed by INBio with the government, the academic sector, and private companies do not need to address the more sensitive issues of indigenous knowledge and direct monetary compensation to communities. Current bioprospecting experience, such as in Costa Rica, indicates that benefits to society at large include an eanced knowledge of the country's biodiversity and improved information dissemination to various sectors of society through guidelines and other publications, information available on the Internet, national and international workshops, and education programs. This scenario is a significant improvement to the informal arrangements that existed prior to the CBD, by means of which individuals would supply germplasm samples to companies without any responsibility or compensation to society.

Perhaps one of the more comprehensive and systematic bioprospecting programs has been the one carried out by the National Cancer Institute (NCI) of the United States. As of 1959 the NCI had major contract-supported programs to collect, identify, and screen plant products and fermentation broth. A modest marine acquisition program was established in 1972. Central objectives of the NCI have been cancer and more recently HIV research. Natural product extracts and other samples are screened *in vitro* against panels of human tumor cell lines and against HIV-infected cells. Acquisition of samples concentrates on marine and unusual microorganisms. Active natural components are isolated by bioassay-guided fractionation and only a handful of them will be pursued to complete structure elucidation. Known natural products are quickly identified via established dereplication.

Benefit sharing agreements go back to the year 1986, when the National Cancer Institute started its second plant acquisition and screening program. Through contracts with the Missouri Botanical Garden, New York Botanical Garden, and the University of Illinois at Chicago, collections in over 25 countries were made under the Letter of Intent and later the Letter of Collection, in which the NCI unequivocally stated its intent to deal with source countries in a fair and equitable manner. By inviting source country scientists to the NCI laboratories, these agreements include a technology transfer component as well as a commitment of the Institute to require successful licensees of their anticancer drugs to enter into further benefit sharing agreements with source country government agencies or organizations. Furthermore, license applicants are required to seek as the first source of supply the natural products from the source country. The letter also acknowledges the achievements and intellectual property rights of local collectors, taxonomists, and healers. In a further development of its benefit sharing policy, for several years the NCI has negotiated Memoranda of Understanding (MOU) with 13 source countries, including Costa Rica. These MOUs are cooperative research agreements and recognize that source country scientists and organizations are committed to performing an increasing share of the operations in-country, as opposed to the export of raw materials.

A successful example of this benefit sharing policy is the company Sarawak Medichem Pharmaceuticals, Inc., which was founded in 1996 as a joint venture between Medichem Research, Inc., a small pharmaceutical company based near Chicago, and the Sarawak state government. Sarawak Medichem is developing (+) calanolide A as an anti-HIV drug. Calanolides are potent inhibitors of retroviral reverse transcriptase and were originally found in leaves and twigs of the tree *Calophyllum lanigerum*, collected in Sarawak, Malaysia, in 1987, through the NCI contract with the University of Illinois at Chicago.

Australia is actively involved in several bioprospecting initiatives and has made significant attempts to add value and advance product development in the country. It has also developed strategic partnerships with Southeast Asian and South Asian countries to increase the availability of germplasm resources. Selected initiatives include those of AMRAD to do bioprospecting on the flora of the Northern Territory and the Australian

Antartic Territory, AIMS focusing on marine samples around the coast of Oz in partnership with several organizations, research on the flora of Queensland by local laboratories linked to ASTRA, and a new laboratory based in Lismore, New South Wales, that will address, among other areas, research and development on medicinal plants.

The International Cooperative Biodiversity Group (ICBG) Program is an experimental program that addresses the interdependent issues of drug discovery, biodiversity conservation, and sustainable economic growth. Funding from the Institutes of the National Institutes of Health (NIH), including the National Cancer Institute, supports awards of approximately $450,000 per year to the five initial initiatives. As of 1997, nearly 4000 species of plants and animals had been examined for biological activity in 13 different therapeutic areas. ICBG-supported projects include acquisition and analysis of natural products derived from biological diversity as potential therapeutic agents for diseases of concern, to both developed and developing countries, like AIDS, malaria, tuberculosis and other infectious diseases, cancers, central nervous system disorders, and heart disease.

The active ICGB projects have made variable progress in knowledge generation, local capacity building, training, and the identification of promising lead compounds. The one in Suriname, briefly reported here, involves the participation of the Virginia Technical Institute, the Missouri Botanical Gardens, Bristol Myers Squibb Pharmaceutical Research Institute, and the National Herbarium of Suriname. After conducting some 14,000 assays from 1993 to 1998 of more than 3300 extracts, Virginia Tech has identified 30 unique extracts that have activity and has isolated 20 chemical compounds that have bioactivity. The most interesting compounds discovered are a group of alkaloids from *Eclipta alba* that have good antifungal activity, better in some cases than the clinically used drug amphotericin B. However, these alkaloids also have weak cytotoxicity, and the decision was thus made not to develop them as antifungal agents.

Meanwhile, Bristol Myers Squibb has put more that 3000 extracts through 32 screens each in six therapeutic areas, with the result that one promising compound (for anticancer activity) is still being tested. While looking for potential pharmaceutical products, project scientists conduct ethnobotanical and random botanical collections using yeast-based assays. The Virginia Tech researchers found there was a slight benefit to the ethnobotanical approaches: 2.8% of the plants collected at random were active in the yeast assays, whereas 3.8% of the shaman-identified plants were bioactive. This small difference is to be expected, given that the assays are not specific to how the shamans use the plants. The ICBG Suriname project has also been able to describe new plant species. This is relevant given that tropical forests have been the source of 60% of the anticancer drugs discovered in the last 10 years and offers a potentially powerful economic reason for preserving the forest, based on the $200 billion market for plant-derived drugs.

The experience of INBio as a participant in one of the initial ICBG projects has been positive in terms of knowledge generation and capacity building, however, it also indicates how the role of the industrial partner was not clearly defined nor negotiated at the outset. This issue has been explicitly raised by Chapela (1996), and in his opinion the five ICBG Consortia reflect the specific interests of the academic partners with little consideration of the objectives of the industrial partners. The interest of industry in the ICBG may be more closely related to discovering microorganisms (which were explicitly excluded) present in some of the samples to be collected (Chapela, 1996).

The isolation of active compounds from tropical plants with a history of medicinal use was the main strategy of Shaman Pharmaceuticals in their efforts to discover and develop novel products for human diseases. The company founded the Healing Forest Conservancy, a nongovernmental organization, explicitly to develop and implement a process by which to return benefits to countries and cultures that chose to contribute to its drug discovery program. Shaman contributed up to 15% of its drug discovery costs to fund projects or programs that were based on the expressed needs of the local communities. Shaman also provided medium-term reciprocity programs that included training for national scientists at Shaman's California laboratories, equipment and resources to conduct research in host countries, and the funding of studies to evaluate sustainable harvesting of indigenous medicinal plants. According to Shaman staff, the practical dilemmas of distributing fair compensation based on intellectual property rights are numerous, and therefore Shaman's approach involved an upfront return and at each further stage of the product development process. The recent and significant reorganization of Shaman in early 1999 will likely affect the ongoing benefit sharing experiences with the communities.

Leading botanical gardens have developed benefit sharing strategies to compensate the countries where botanical specimens are initially collected. Through their collaboration with the National Cancer Institute, the New York Botanical Garden and the Missouri Botanical Garden voluntarily adopted policies on the research of natural products in the mid-1980s. Most notably the Missouri Botanical Garden committed itself to enter into commercial research agreements only with the provision that royalties will be paid to the source country and that the Garden itself will not receive any percentage of such royalties. In 1981 the New York Botanical Garden established the Institute of Economic Botany to focus a portion of the Garden's research enterprise on applied botanical questions of great human concern. One of its forest projects is the Belize ethnobotany project, begun in 1988, to conduct an inventory of the diversity of medicinal plants and to preserve this richness. In Europe, the Royal Botanical Gardens at Kew took the lead and its benefit sharing policy also includes those materials that had been collected prior to the ratification of the CBD.

A very innovative and comprehensive strategy is that of South Africa, where traditional healers, the government, and private and academic sectors are joining efforts to undertake bioprospecting. The initial investment (40% government and 60% private sector) of U.S. $10 million will fund the investigation of 23,000 plant species plus marine organisms and microorganisms. The plan is to develop international partnerships, however, most of the R&D will be done in-country.

III. CONCEPTS AND PRACTICES OF BIOPROSPECTING: THE CASE OF COSTA RICA

A. Nature-Based Initiatives and the Role of the National Biodiversity Institute

Costa Rica has developed a system of conservation areas in which over 25% of the national territory is allocated to nature reserves and national parks. Both the government and civil society have a role in conservation and have contributed to creating awareness of the importance of biodiversity at the national and international levels. In terms of economic benefits, Costa Rica derives its main source of income from nature-based tourism; this sector yielded over U.S. $850 million in 1998 and appears to be growing steadily. Surveys indicated that 70–75% of all tourists come to experience nature. This is an impressive development when compared to the traditional economic sectors of banana and coffee production $500 million and $350 million, respectively).

Other relevant initiatives to value biodiversity include payment for environmental services such as CO_2 fixation. A 5% fuel tax yields approximately $40 million yearly, which is invested in carbon sequestration programs to support the conservation of forests and reforestation. Serious consideration is also being given to charge users for the actual costs of water and electricity, and initial experiments are under way. Currently, water bills reflect essentially distribution costs and not the value of this essential commodity to society.

Costa Rica is considered to be a megadiversity country. Most of its territory has a diversity density of over 5000 species of vascular plants per 10,000 km^2. Only 10 other countries on earth (Nicaragua, Panama, Columbia, Ecuador, Peru, Bolivia, Brazil, China, Malaysia, and Papua New Guinea) have areas with similar diversity (Barthlott *et al.*, 1996). However, of the total number of species estimated for the country, only a minor portion has been described. Vertebrates and vascular plants have been well studied, with over 90% of the species documented and characterized. For all other groups of organisms, including insects, algae, fungi, and bacteria, there are enormous gaps of knowledge.

The management and development of biological resources fall under the responsibility of the Ministry of Environment and Energy (MINAE), specifically the National System of Conservation Areas (SINAC), which is the institution in charge of conserving and promoting the sustainable use of the country's natural resources. The eleven Conservation Areas in which the country has been divided constitute the various regions established by MINAE to carry out a decentralized management of biodiversity and natural resources, with the participation of the communities surrounding the protected areas.

Executive Presidential Decree No. 19153 of June 5, 1989, established a Biodiversity Planning Commission, with representatives from various governmental organizations, higher education institutions, and nongovernmental conservationist organizations. This Commission recommended the creation of the National Biodiversity Institute (INBio) as a nonprofit, nongovernmental organization to serve the public good. INBio was legally registered on October 26, 1989, as a civil association governed by an Assembly of Founders and a Board of Directors.

INBio works under the premise that a tropical country will be able to conserve a major portion of its wild biodiversity if this biodiversity generates enough intellectual and economic benefits to make up for its maintenance. INBio and the Ministry of Environment and Energy established a collaborative agreement that allows INBio, within the existing legal framework, to conduct biological inventories, biodiversity prospecting, and management and distribution of Costa Rican biodiversity information.

One of INBio's goals is to generate greater awareness of the value of biodiversity and, thereby, promote its conservation and sustainable utilization based on the demands of national and international users. This goal is fulfilled through the integration of the following processes:

- biodiversity inventory focused particularly on the country's protected wildlands;
- conservation for development;
- identification and promotion of sustainable uses of biological resources;
- organization and management of information; and
- generation and transfer of knowledge.

The Biodiversity Inventory Program is carried out in Conservation Areas by "parataxonomists." They are in charge of collecting, preparing, and mounting the biological material with pertinent field information, and local curators are responsible for the taxonomic classification of such material, with support from national and international experts.

The Conservation for Development Program helps strengthen the capacity of INBio and assists in the tasks assigned by the National Conservation Strategy, which is based on protecting, understanding, and using biodiversity in a sustainable manner.

The Biodiversity Information Management Program processes and organizes the information generated by the inventory and other national collections under various formats, according to the needs of users. Thus, it seeks to process and integrate texts, maps, photographs, and quantitative information about Costa Rican biodiversity.

Natural resources constitute an essential source for new medicines, agricultural products, coloring compounds, resins, and other uses. Through the Bioprospecting Program and in partnership with national and international organizations, INBio carries out a systematic search of chemical compounds, genes, and micro- and macroorganisms that may help find new products for the benefit of humanity.

The need to disseminate biodiversity information motivated the creation of the Social Outreach Program as a means to distribute information to different sectors of society. It includes production and publication of printed and audiovisual material on biodiversity issues for various users, special programs for elementary and high school students, and information for visitors. This program also cooperates with various national commissions of private and public organizations.

B. Partnerships and Management Issues

Current bioprospecting approaches require a significant degree of partnership at various levels. In Costa Rica, INBio has developed a close alliance with the government (Ministry of Natural Resources and Energy) that makes bioprospecting development possible in the nation's Conservation Areas. This first level of partnership is essential and is based on trust, understanding, and clear-cut responsibilities and benefits (see Section III,D).

Another essential level of partnership involves the academic sector. Universities, both local and international, complement the required expertise in biology, chemistry, microbiology, and biotechnology that is central to the concept and need of adding value at the national level. Last but not least, a strong alliance with the private sector (conservation, pharmaceutical, agricultural, biotechnology, and fragrance companies) allows market forces and the needs of society to determine the products and services to be pursued through bioprospecting. The country has a limited capacity to invest in science and technology, and therefore partnerships with the academic and private sectors are the only way to accomplish many of the stated goals.

C. Mechanisms and the Legal Framework

Managing the complexity of the institutional interactions and addressing the needs and expectations of the various partners require careful planning and organization. Costa Rica and INBio cannot provide a prescription for bioprospecting success, but do offer 10 years of experience full of drawbacks, challenges, and achievements.

The requirements of a bioprospecting program are relatively simple to state: vision, leadership, negotiation, planning, understanding partners' strengths and expectations, teamwork, realistic research plans, and sufficient local expertise in science and informatics. But these elements are not necessarily simple to operationalize. It is widely perceived that being a small country and having a well-educated workforce and the right incentives make the process easier in a country like Costa Rica. All this is probably true, however, there are few comparisons available to ascertain the validity of this hypothesis.

TABLE III

The Legal Framework for Biodiversity Conservation and Bioprospecting in Costa Rica

- Forestry Law, 1995
- National Parks Law, 1977
- Wildlife Conservation Law, 1992
- Ministry of Environment Law, 1990
- Convention on Biological Diversity (CDB), 1993
- General Environmental Law, 1995
- Biodiversity Law, 1998

A country's legal framework, based on both theory and experience, sets the fundamental organization and incentives not only for conservation but also for bioprospecting. Table III highlights significant pieces of legislation that have made conservation and utilization of biodiversity possible in Costa Rica. The 1998 Biodiversity Law, a law of consensus among the various stakeholders, reflects in a significant way the country's prior experiences in conservation and bioprospecting.

D. Criteria

Out of the 45 articles contained in the Convention on Biological Diversity (CBD), INBio and the National System of Conservation Area (SINAC) of the Ministry of Environment and Energy have emphasized the following ones in the design, approval, and implementation of bioprospecting agreements.

1. Access

Signatory countries of the CBD have agreed to facilitate access to other members. INBio and SINAC comply with this principle provided that other agreed criteria (presented in the following sections) are an integral part of the access agreement.

2. Compensation

In the INBio/SINAC strategy, compensation is negotiated at two different levels:

(a) All research costs related to a bioprospecting agreement (e.g., collecting, GIS, natural history extraction, determination of activity, etc.) are paid by the interested company. A contribution equaling 10% of all operational costs is transferred to SINAC for conservation purposes.

(b) In case a product based on a sample provided by Costa Rica reaches the market, royalties are payable in lieu of patent rights. Current patent laws in Costa Rica offer limited protection and are therefore unsuitable for products destined for the world market. Royalties vary significantly, depending mostly on added value generated within the country.

Literature surveys indicate that 1–15% royalty ranges are the norm; INBio's experience coincides in most cases with these figures. The transfer of resources from bioprospecting agreements to research and conservation programs is shown in Table IV, and the list of active bioprospecting agreements (both academic and with industry) is given in Table V.

3. Transfer of Technology

Access to partner companies' research processes and bioassays of interest to INBio and its local partners is negotiated as part of the transfer of technology. INBio's current research capacity in bioprospecting is largely based on technology acquired through such agreements with the private and academic sectors.

4. Training

Contributions that a host country can make in bioprospecting research and product development are essential

TABLE IV

Contributions and Payments from Bioprospecting Agreements (U.S. $)

	1993[a]	1994	1995	1996	1997	1998 (Jan–Feb)	Total
Ministry of the Environment and Energy (MINAE)	110,040	43,400	66,670	51,092	95,196	24,160	390,558
Conservation Areas	86,102	203,135	153,555	192,035	126,243	29,579	790,649
Public Universities	460,409	126,006	46,962	31,265	34,694	326,486	1,025,822
Other groups at INBio	228,161	92,830	118,292	172,591	129,008	0	740,882
Total	884,712	465,371	385,479	446,983	385,141	380,225	2,947,911

[a] Estimated amounts since 1991.

Source: INBio, 1998.

TABLE V

INBio Bioprospecting Agreements

Partner	Year started	Application
University of Costa Rica	1991	General agreement
	1998	Operation of an NMR unit
University of Strathclyde (U.K.)	1991	New pharmaceutical products from plant sources
Merck & Co. (New Jersey, USA)	1991/1994/1996/1998	Pharmaceutical and veterinary products from plants and microbial sources
BTG/ECOS-Costa Rica	1992	Development of a bionematicide from *Lonchocarpus* sp.
Cornell University	1993	Drug discovery
Universidad Nacional, Costa Rica	1993	General agreement
Bristol Myers Squibb/Cornell University	1994	Insects as source of new compounds for pharmaceutical industry
Instituto Tecnológico de Costa Rica	1994	General agreement
Givaudan Roure (New Jersey, USA)	1995	Fragrances and aromas from Costa Rican biodiversity
University of Massachussets	1995	New insecticides
Diversa (San Diego, USA)	1995/1998	New enzymes from extremophilic and other microorganisms
Government of Canada	1996	Debt for nature swap
INDENA (Italy)	1996	New antimicrobials for dermatological use from plant sources
EARTH/NASA/Other Latin American institutions	1997	Chaga Space Project: compounds against regulatory enzymes
University of Strathclyde (U.K.)	1997	Pharmaceutical products
Office for Scientific, Technical and Cultural Affairs (Belgium)	1998	MOSAICC Project
Phytera (Massachusetts, USA)	1998	Pharmaceutical products from plant sources

elements of any added-value strategy. In every agreement, one or two training sessions for Costa Rican scientists and technicians per year (normally of short duration) are negotiated at the company's expense.

5. Sustainable Uses

INBio and its partners will not engage in extractive strategies should a forest resource become of interest to the market. Instead, they will pursue chemical synthesis approaches if possible, a domestication strategy if feasible and desirable. In the latter case, a species may be successfully domesticated and become a new agricultural crop or a new productive enterprise with the potential to generate economic development.

E. Lessons Learned, Challenges, and Opportunities

Costa Rica and INBio have gained considerably from bioprospecting agreements, through knowledge about the national biodiversity, transfer of technology training, and development of negotiating capabilities. However, we have also learned that bioprospecting is not the gold mine that was initially envisioned. There are no guarantees that royalties will materialize in the future and therefore they should not be included in decision-making and budgetary processing. Obviously, if they become a reality they will be a welcome incentive and support for conservation and research.

We have learned that partnerships and alliances are indispensable, that innovation must be at the forefront of our activities, and that controversy and misunderstandings are part of the daily routine. We have also learned, in order for history not to repeat itself, that Costa Rica must develop internal R&D capacity leading to advanced product development processes and in certain cases to finished products. Private companies have shown a cyclical interest in natural products, and therefore bioprospecting has had and will continue to experience peaks and valleys of demand.

At INBio, bioprospecting is only one component of a much wider national biodiversity strategy. The major factors responsible for INBio's initial accomplishments have been solid political support, leadership and vision, a commitment to innovation, a diversified portfolio of goals, and strategic alliances (Mateo, 1996). Early bioprospecting efforts emphasized demand-driven agreements with pharmaceutical, biotechnology, agricultural, and fragrance companies. This line of activity will certainly be maintained and enhanced, and it should evolve toward higher levels of research and services that are performed within Costa Rica (the added-value contribution). In parallel, and given the particular conditions in Costa Rica, INBio will also be involved in the development of new biodiversity products (nutraceuticals, phytomedicines, ornamentals, etc.) with national industries and local communities. This strategy will involve a pilot product development phase followed by marketing of biodiversity-derived products. Cost recovery mechanisms will be designed to support new product development initiatives.

The recently born concept of modern bioprospecting is already being challenged by new developments in science and technology. In an endeavor to reduce risks and to increase the probability of success, companies are trying to substitute natural (incalculable) products with artificial (calculable) products wherever possible. The success rate in drug discovery is estimated to be about one product out of 10,000 screened chemicals or plant species for one field of diseases. To maintain 10% annual growth, or roughly \$1 billion in new revenues, drug companies must bring two or three new products to the market each year, which is four to six times the current average achieved by the top producers. To speed up the discovery process, companies are investing in Ultrahigh Throughput Screening (UHTS) technologies, in which up to 100,000 compounds can be exposed each day to a growing array of bioassays. The number of extracts available from biological resources can scarcely satisfy this machinery and are now challenged by the chemical diversity provided by combinatorial chemistry, which synthesizes all possible combinations of compounds from various sets of chemical building blocks.

It is too early to judge whether combinatorial chemistry will be as successful as natural sources in the generation of new drug candidates. Many scientists doubt that novel lead structures will result from these combinatorial efforts, which can compete with those evolved in nature during an evolutionary process of millions of years. However, the technology can be a powerful tool to complement natural sources by generating millions of derivatives from natural leads. It should be noted that many highly successful drugs, which have been introduced to the market only recently, are derivatives of natural leads that were discovered many decades ago. Examples are diclofenac (from salicin), clarithromycin (from erythromycin), and ceftriaxon (from cephalosporin).

In conclusion, the hunters, gatherers, and agriculturalists of the Old and New Worlds are, in some ways, the equivalent of the modern biochemical and gene bioprospectors. Perhaps the methods and strategies differ, yet the goals are essentially the same: survival, well-being, and spiritual and material advancement.

See Also the Following Articles

AGRICULTURE, TRADITIONAL • BREEDING OF PLANTS • ECOLOGICAL GENETICS • ECONOMIC VALUE OF BIODIVERSITY • GENE BANKS • PLANT SOURCES OF DRUGS AND CHEMICALS

Bibliography

Barthlott, W., Lauer, W., and Placke, A. (1996). Global distribution of species diversity in vascular plants: Towards a world map of phytodiversity. *Erdkunde* **50**, 317–328.

Bills, G. F., and Polishook, J. D. (1994a). Abundance and diversity of microfungi in leaf litter of a lowland rain forest in Costa Rica. *Mycologia* **86**, 187–198.

Bills, G. F., and Polishook, J. D. (1994b). Microfungi from decaying leaves of *Heliconia mariae*. *Brenesia* 41–42, 27–43.

Chapela, I. H. (1996). La bioprospección en la era de la información: Un análisis crítico de las iniciativas de conservación asociadas con el descubrimiento de nuevos fármacos. In *Biodiversidad, Biotecnología y Desarrollo Sostenible en Salud y Agricultura: Conexiones Emergentes*, pp. 29–49. Organización Panamericana de la Salud, Washington, D.C.

Constanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., and van der Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature* **387**, 253–260.

Crouse, G. D. (1998). Pesticide leads from nature. *Chemtech (American Chemical Society)*, Nov. 98, 36–45.

Ehrlich, P. R., and Raven, P. H. (1964). Butterflies and plants: A study in co-evolution. *Evolution*, **18**, 586–608.

Grünwald, J. (1995). The European phytomedicines market: Figures, trends, analyses. *HerbalGram* **34**, 60–65.

Hill, B. (1997). *Innovationsquelle Natur*. Shaker Verlag, Aachen, Germany.

Hill, B., Corrales, J., and Nader, W. F. (in press). *Biodivesität und Tropenbionik. Biologie in unserer Zeit.*

Janzen, D. H. (1975). *Ecology of Plants in the Tropics*. Edward Arnold, London.

Madigan, M. T., and Marrs, B. I. (1997). Extremophiles. *Sci. Am.* **276**, 82–87.

Mateo, N. (1996). Wild biodiversity: The last frontier? The case of Costa Rica. In *The Globalization of Science: The Place of Agricultural Research* (C. B. Friedheim and K. Sheridan, eds.), pp. 73–82.

International Service for National Agricultural Research (ISNAR), The Hague, Netherlands.

Nader, W. F., and Hill, B. (1999). *Der Schatz im Tropenwald—Biodiversität als Inspirations- und Innovationsquelle*. Shaker Verlag, Aachen, Germany.

Piperno, D. R., and Pearsall, D. M. (1998). *The Origins of Agriculture in the Lowland Neotropics*. Academic Press, San Diego, California.

Reid, W. V., Laird, S., Meyer, C. A., Gámez, R., Sittenfeld, A., Janzen, D. H., Gollin, M. A., and Juma, C. (1993). *Biodiversity Prospecting: Using Genetic Resources for Sustainable Development*. World Resources Institute, Washington, D.C.

Tamayo, G., Nader, W. F., and Sittenfeld, A. (1997). Biodiversity for bioindustries. In *Biotechnology and Plant Genetic Resources* (J. A. Callow, B. V. Ford-Lloyd, and H. J. Newbury, eds.), pp. 255–280. CAB International, Wallingford, United Kingdom.

ten Kate, K. (1995). *Biopiracy or Green Petroleum? Expectations and Best Practice in Bioprospecting*. Overseas Development Administration, London.

Thayer, A. M. (1998). Pharmaceuticals, redesigning R&D. *Chem. Engineering News*, February 23, 25–37.

Wink, M. (1993). Production and application of phytochemicals from an agricultural perspective. In *Phytochemistry and Agriculture* (T. A. Beek and H. Breteler, eds.), pp. 171–213. Clarendon Press, Oxford, United Kingdom.

Wyvratt, M. J. (1988). Evolution of angiotensin-converting enzyme inhibitors. *Clin. Physiol. Biochem.* **6** (3–4), 217–229.

BIRDS, BIODIVERSITY OF

Jeremy J. D. Greenwood
British Trust for Ornithology

GLOSSARY

biomass The total dry mass of a population or community of animals or plants.

congener A species that belongs to the same genus as the one under discussion.

elfin woodland Small, stunted tree growth characteristic of forests at higher elevations in warm, moist regions.

endothermic An animal whose body temperature is largely determined by its own metabolism rather than by the temperature of the environment.

epiphyte A plant that uses another plant, usually a tree, for support but not for nourishment.

niche The unique position occupied by a particular species in terms of the function that it performs within the community.

polyandrous A female that has more than one (male) mate.

polygynous A male that has more than one (female) mate.

trophic Relating to feeding.

THE LIFE OF BIRDS is based on flight, which has both constrained their diversification and opened ecological opportunities. It affects their size and their abundance. Their ecology is based on the exploitation of patchy and variable resources, migration being the most obvious manifestation of this. A few birds are flightless. The taxonomic diversity of birds is limited. Across all continents, bird species diversity is greater in wooded than in forested habitats, generally being related to habitat complexity and productivity. Coexisting species eat different foods and often use different parts of the habitat. Bird species diversity is greater where there is greater habitat diversity; it is less at higher altitudes; and it is generally less as one moves from the tropics toward the poles, with strong local modifications. It may be correlated with that of other animals, but not always. Bird data have been used to test hypotheses about the causes of the latitudinal gradient in biodiversity. There is evidence that it is partly dependent on tropical regions providing more ecological opportunities, more habitable area, and more utilizable energy. No species is found all over the world. Some closely related species replace each other geographically. Some parts of the

world, mainly tropical, have unusually large numbers of endemic, restricted-range species. Geological and evolutionary history has resulted in the avifaunas of different regions of the world differing in richness and taxonomic origin. Unrelated species living in different places but similar habitats have often undergone evolutionary convergence. Birds are good colonists of islands, where they subsequently evolve in the presence of few (if any) other land vertebrates. More species are found on larger islands, on those that once formed parts of larger landmasses, and on those closer to sources of colonists. Islands that become cut off from the mainland or which are fragmented lose species because extinctions are not then sufficiently matched by colonizations. Humans have caused the extinction of many birds, in both historical and prehistorical times, especially on islands. The chief current threats to them are habitat loss, hunting, agricultural intensification, and the introduction of alien species to islands. Reducing these threats will be difficult. Avian biodiversity continues to be lost both by large-scale declines of many species in some parts of the world and by the introduction of species without their native ranges, which diminishes the differences in the bird communities of different lands.

I. FLIGHT: CONSTRAINTS AND OPPORTUNITIES

A. Birds as Flying Machines

The feather is one of the most remarkable structures ever to have evolved. It has allowed birds to become highly effective flying machines and to exploit resources in ways impossible for other vertebrates, although it has also placed constraints on them, considerably restricting their morphological diversification and thus limiting the ecological opportunities open to them.

Feathers provide highly effective insulation. Simply stated, the downy bases of the contour feathers trap a layer of air close to the body, and the integrity of this layer is preserved by the vanes of the feathers, which form a largely windproof and waterproof covering, like the tiles on a roof. Weight for weight, birds' plumage is considerably warmer and more waterproof than the hair and wool of mammals. Like mammals, therefore, birds are enabled to be endothermic, usually maintaining body temperatures well above their surroundings and thus leading active and independent lives.

The structure of feathers allows them to form strong but lightweight aerofoils, thus facilitating the evolution of wings. Since the wings are made up of many feathers, they are very adjustable, allowing effective and maneuverable flight. Because feathers have no blood supply, birds do not lose inordinate quantities of heat through their wings, as bats are in danger of doing. The bat wing is difficult to repair; in contrast, a damaged feather is renewed at the next molt. The strength of feathers allows birds to have long tails, which are used for both aerodynamics and social signaling, even though the bony and fleshy tail on which those feathers are mounted is reduced to a short stub.

The shape of wings varies considerably. Long, pointed wings provide high speed but little maneuverability and lift; they are characteristic of birds that need to fly fast or migrate long distances and which live in open places, such as plovers *Pluvialis.* Short, rounded wings give maneuverability and lift but little speed; they are characteristic of woodland birds such as tits *Parus.* Large, broad wings allow birds to soar on rising air, covering long distances while patrolling the landscape for food; they are characteristic of large hawks, eagles, and vultures. Long, straight wings allow birds to glide on ocean winds; they are characteristic of albatrosses and other such seabirds.

Most vertebrates tend to have long, flexible bodies. Not so with birds: Flight requires a short, rigid trunk. This is achieved not only through the trunk being short and broad but also through the locking together of the skeletal elements, providing a firm base for the operation of the flight muscles while keeping skeletal weight down. Long bones are hollow, reducing weight. The compactness of the body is increased by having the major muscles of both wings and legs close to the center of mass, rendering locomotion mechanically efficient. Because the trunk is rigid, the neck must be long so that the head can be moved around. (The length of the neck is generally not apparent because of the covering of feathers.) Were birds to have a tidal airflow ventilating the lungs, this would be a major disadvantage, because in such a system there is a "dead space" lying between mouth and lungs into which air is drawn but never used. Birds have a through-flow so that all inspired air passes through the lungs. This is why they do not need to breathe as often as mammals of comparable size, even though they operate at the high metabolic rate that flight demands. Birds' hearts are larger than those of similarly sized mammals to underpin their rapid metabolism.

The long neck means that it is particularly important for the head to be light. Modern birds are toothless and do not have the heavy jaws typical of many reptiles and mammals. Although their beaks and tongues are capable of remarkable feats of manipulation (such as finches husking seeds in seconds), this means that food is largely swallowed without chewing. To compensate, most species have part of the stomach modified as a muscular gizzard within which the food is ground. Indigestible elements of the food, such as bones, hair, arthropod exoskeletons, and husks, tend to be regurgitated in pellets.

As bats demonstrate, flight does not absolutely preclude viviparity; however, birds are the only vertebrate class in which no species bears live young. Weight restrictions result in most bats producing only one young per year, which prevents them from adopting "r-selected" lifestyles characterized by rapid reproductive rates. Some birds, in contrast, may produce well over 10 young each year. In those many species in which the young remain in the nest until fully grown, the parents' ability to fly allows them nonetheless to forage over comparatively large areas.

Being so mobile in three dimensions, birds use vision and hearing as their dominant senses; they communicate mainly by sight and sound.

B. Size and Abundance

The constraints imposed by flight are nowhere as apparent as in the upper limit of bird size. The lifting power of wings depends on their surface area and on the cross-sectional area of the muscles powering them; however, the weight to be lifted depends on the volume of the body, so the ratio of weight to power increases with the size of the animal, imposing an upper limit on the size of flying animals. This appears to be approximately 15 kg for modern birds, exemplified by some swans, pelicans, and bustards, although some extinct species seem to have been appreciably larger. Flightless birds may be larger—more than 100 kg in the case of the ostrich *Struthio camelus* and approximately 450 kg in the extinct elephant bird *Aepyornis maximus;* however, even these are dwarfed by African elephants *Loxodonta africana* at 7 tons and blue whales *Balaenoptera musculus* at 100 tonnes. Flightless birds have no particular advantage over mammals and there are few of them.

Metabolic rate seems to impose similar lower limits on the size of birds and mammals: Kitti's hog-nosed bat is 1.5–2.0 g, shrews such as the pygmy white-toothed shrew *Suncus etruscus* 2–3 g, and the bee hummingbird *Calypte helenae* 1.6–3.0 g. The problem is that metabolic rates of small endotherms are disproportionately great, partly because their large surface to volume ratio results in disproportionate heat loss. Not only is it difficult for them to ingest enough food to maintain such high rates but also there is an upper limit to the rates imposed by the rate at which their hearts can beat. The 1200–1400 contractions per minute observed in the smallest species may be the maximum achievable.

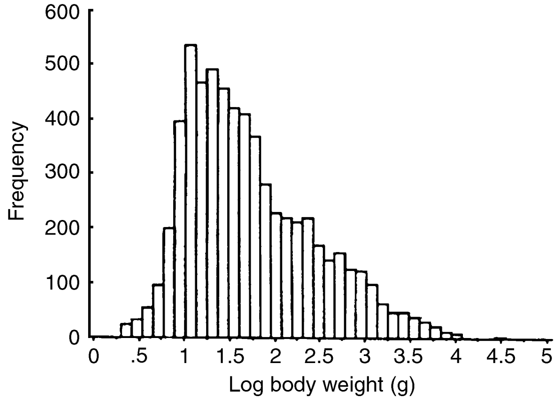


FIGURE 1 The frequency distribution of the logarithm of body mass of the 6209 species of birds for which data are available (reproduced with permission from Blackburn and Gaston, 1994).

Most birds are small—half of the species weigh less than 38 g (Fig. 1). They are larger toward the poles (Fig. 2), perhaps because it is easier for large birds to keep warm or because they are more able to migrate long distances. Within the same region, smaller birds are more common in wooded than in open habitats; the former give them better protection from weather and predators as well as provide many foraging opportunities among twigs and foliage, which heavy birds can-

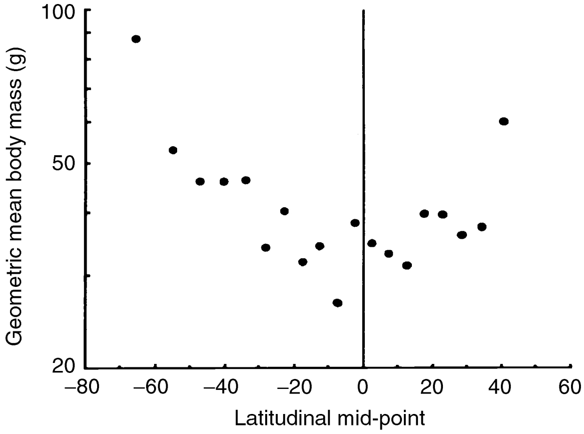


FIGURE 2 The relationship in New World birds between body mass (geometric mean) and latitude (negative values are in the Northern Hemisphere) (reproduced with permission from Blackburn and Gaston, 1996a).

not exploit. Waterbirds tend to be large to aid heat conservation.

In many birds, males are slightly larger than females; they are particularly large in many polygynous species, which suggests that the reason for the difference is the advantage that larger males have in competing for females. Females tend to be larger in polyandrous species and also in species in which males perform aerobatic displays as part of courtship.

On a logarithmic scale, the abundances of bird species in a region are approximately normally distributed, with some skew because of a tail of very scarce breeders (Fig. 3). The same is probably true on a world scale. The Seychelles warbler *Acrocephalus sechellensis* was once confined to Cousin Island (29 ha), its population reaching a low of 50 pairs. Numerous other species confined to small, remote islands may have maintained populations of no more than a few hundred throughout their history. Various species have total world populations much less than this but usually as a result of human interference. Wilson's petrel *Oceanites oceanicus* is probably the most common seabird, at 100 million or more. On land, the red-billed dioch *Quelea quelea,* of which individual flocks may number over 1 million, is almost certainly even more numerous; house sparrow *Passer domesticus* and starling *Sturnus vulgaris* are so widespread, partly through human introduction, that they may have total populations even greater than those of *Q. quelea,* which is confined to the more arid parts of sub-Saharan Africa.

At both local and regional scales, smaller birds are more abundant than larger ones (Fig. 4). Species that are more abundant in places where they occur occupy more sites within regions (Fig. 5) and have more extensive geographical ranges (Fig. 6), though on both scales colonial species (which occupy few sites but at high densities) weaken the relationship. The correlation tends to be weaker outside the breeding season, perhaps because many noncolonial species then form flocks. Various reasons for this relationship have been suggested, but the evidence favors none in particular.

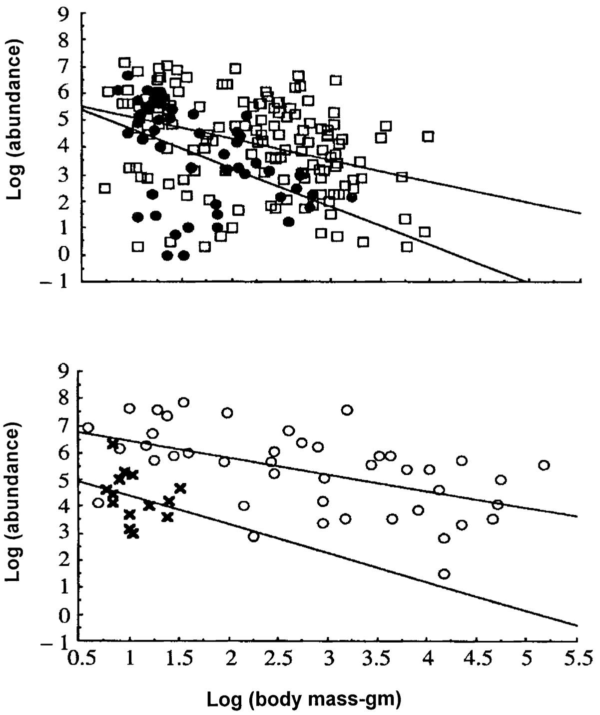


FIGURE 4 Relationships between abundance and body mass in nonflying mammals (○), bats (crosses), and resident (□) and migrant (●) nonmarine birds breeding in Great Britain. Note that both scales are logarithmic so that the mean difference in abundance between birds and nonflying mammals of similar size is approximately 45-fold (redrawn from Greenwood *et al.,* 1996).

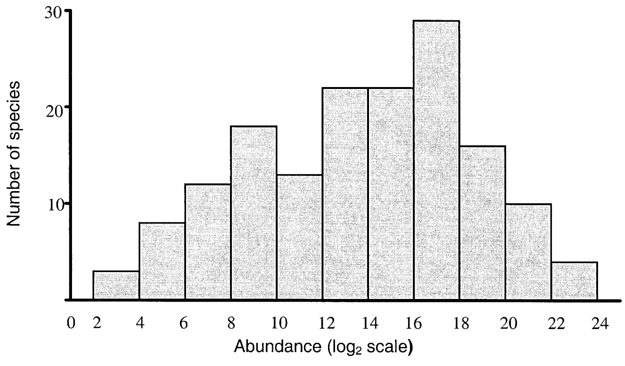


FIGURE 3 The frequency distribution of the population sizes of the 157 nonmarine native birds regularly breeding in Britain (British Trust for Ornithology data).

In Britain at least, summer visitors tend to be smaller, on average, than residents, perhaps reflecting difficulties that small birds have in keeping warm in winter. They are substantially less abundant than residents of the same size (Fig. 4). (In North America, summer visitors are more abundant than residents; this is also true in Europe if one does not consider size.)

C. The Ecological Diversity of Birds

A 1-kg mammal can travel at sustained speeds of 6 km/hr, expending 118 kJ/km in doing so. A bird of the same size can travel at 47 km/hr, expending only 49 kJ/km. The power of flight thus allows birds to cover larger areas in the courses of their lives and this is the

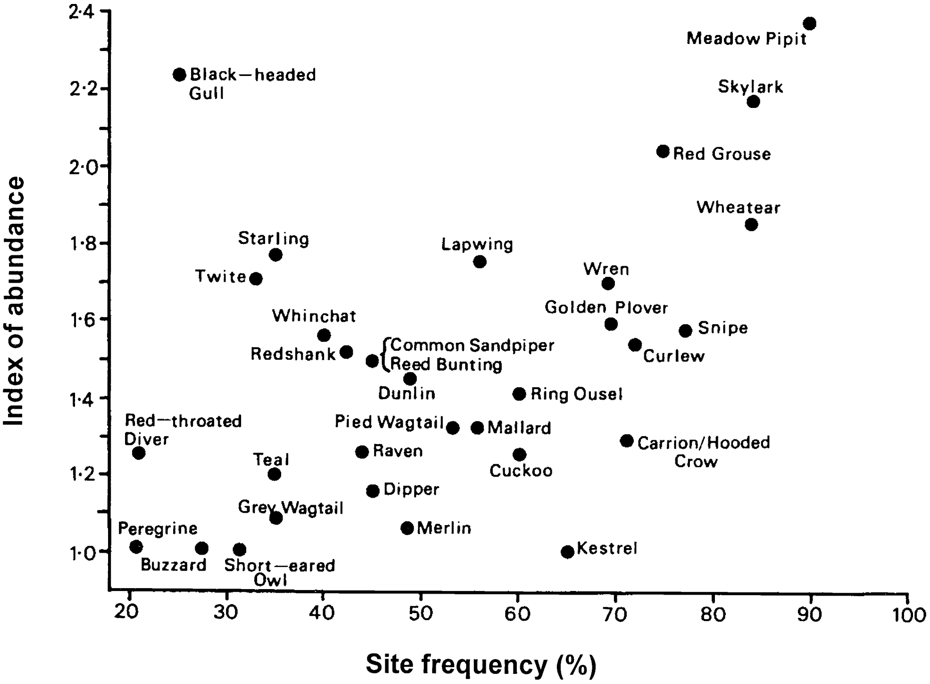


FIGURE 5 The relationship between the abundance of British upland birds in the sites at which they do occur and their frequency of occurrence at sites (reproduced from Fuller, 1982, with permission of the British Trust for Ornithology).

key to their ecology: They tend to use scattered and variable resources much more than do earth-bound animals. As a result, they are able to occupy significantly narrower ecological niches, subdividing resources more finely. This may be the reason why, within the same geographical area, the abundance of individual species of nonflying mammals is very much greater than that of birds of the same size (Fig. 4). This, plus the fact that the average mammal is several times larger than the average bird, means that birds make up much less of the total biomass of terrestrial animals than do mammals (e.g., 13,000 tons for British resident birds versus 158,000 tons for mammals).

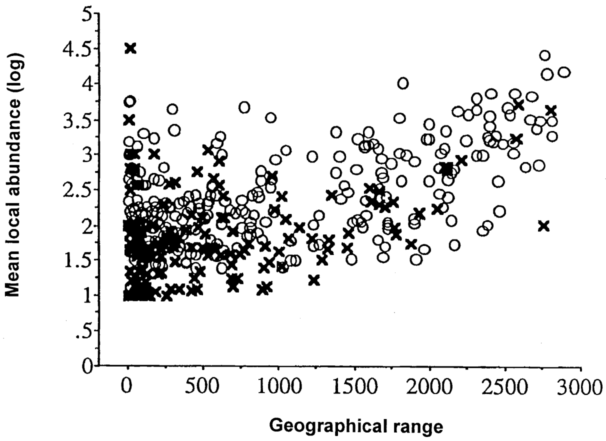


FIGURE 6 The relationship of mean local abundance and range size in European nonmarine breeding birds. The crosses mark species of special conservation concern (reproduced with permission from Gregory *et al.*, 1999).

In contrast to such local restrictions on numbers, birds are able to occupy large geographical ranges because their powers of flight allow them to discover areas that provide them with the resources they require. Thus, local areas tend to be occupied by a higher percentage of the whole continental species pool of birds than is the case for other well-studied groups, such as mammals, amphibians, lizards, and butterflies.

Many waterbirds make seasonal use of water bodies that are impermanent or which freeze over in winter. On a shorter timescale, some species of birds (but few nonflying animals) move onto seashores in large numbers as tides recede, exploiting resources that are often rich but which are only briefly available. Pelagic seabirds use the efficiency of flight in another way, covering large areas in search of food that may be constantly available but is sparsely distributed. Outside the breeding season, they may move over thousands of kilometers, often along definite tracks determined by winds, currents, and seasonal availability of food. Seabirds quickly discover shoals of fish that are close to the surface, such as at upwelling fronts of cold water, which shift unpredictably. Even those land birds that are resident in a small area are constantly moving about that area in ways that allow them to subsist on food that is either sparsely distributed (such as seeds in temperate grasslands in winter) or which occurs in patches, especially temporary patches (such as carcasses, fruiting trees, nectar-bearing flowers, ant and termite swarms, insects disturbed by ungulate herbivores, and small animals disturbed by fire). Outside the breeding season they may temporarily use habitat patches that are inadequate to support them while breeding.

Birds are able to exploit aquatic habitats not just because of their powers of flight but also because of the effectiveness of their plumage for insulation. This is particularly important for smaller animals (because of their high surface to volume ratio) and it is noteworthy that there are few aquatic mammals weighing less than several kilograms. Those that do must frequently leave the water to dry out their fur. Insulation is less important for larger animals, and terrestrial mammals have given rise to cetaceans, sirenians, and seals, apparently excluding birds from occupying these ecological zones.

The Australian central desert, lying between two rainfall belts, may experience occasional rain at any time, anywhere; approximately 30% of Australian birds are essentially nomadic, settling to breed quickly in places where rain has fallen. Although some desert birds can survive without drinking much free water, others cannot; rather, they are able to live in deserts by using their powers of flight to reach isolated drinking places. Sandgrouse (Pteroclididae) may make round-trips of over 100 km daily between their nest sites and water holes, with males transporting water back to the chicks that is soaked into the modified breast feathers.

Although flight allows birds to exploit patchy resources, it also restricts what they feed on. To keep weight down, birds have short guts, through which food passes quickly; therefore, they tend to feed on particularly nutrient-rich items, generally eschewing food that needs to be taken in bulk. Insects (and other terrestrial invertebrates when available), littoral invertebrates, fruit, seeds, fish, and other small vertebrates are taken by many species but, despite leaves making up such a large proportion of the biomass on land, comparatively few birds are specialist leaf eaters (Table I). Some grouse, geese and some ducks, some pigeons, the New Zealand ground parrot *Strigops habroptilus,* and the hoatzin *Opisthocomus hoazin* are exceptions: Most tend to feed selectively on the most nutritious leaves and to move their food through the gut rapidly.

Carrion is very sparsely distributed; therefore, only a few birds specialize on it, although there has been convergent evolution on this specialism in New and Old World vultures. Similarly, nectar is only found in sufficient quantities for birds to use in some parts of the world and efficient nectar-feeding requires anatomical and behavioral specializations so there are few families for which nectar is the primary food (and, as with the vultures, there has been a convergence of New and Old World forms: humming birds and sunbirds, respectively).

The various diets of birds are associated with a variety of bill shapes (Fig. 7). There are subtler variations: Frugivores with wider gapes take larger fruits, and finches with deeper bills take larger and harder seeds. Even within species, differences between individuals in bill size may be reflected in diet, the most extreme example being the African finch *Pyrenestes ostrinus* in which there are two genetically distinct forms specializing on sedge seeds that differ in hardness. The extinct Huia *Heteralocha acutirostris* of New Zealand fed on insects in dead wood, with the female probing into

TABLE I

Numbers of Bird Families Feeding on Different Food Categories[a]

Food category	Primarily	Regularly	Sparingly	Infrequently, if at all
Green plants, buds	2	13	13	140
Seeds	4	31	24	109
Fruit	19	34	21	94
Nectar	3	5	7	152
Insects	50	58	32	28
Other terrestrial invertebrates	2	10	23	133
Littoral and benthic invertebrates	6	10	13	139
Small vertebrates (<5 kg)	4	14	29	121
Large vertebrates (5 kg +)	0	0	3	165
Fish, crustaceans, squid, etc.	20	10	11	127
Carrion	1	4	6	157

[a] From Morse (1975).

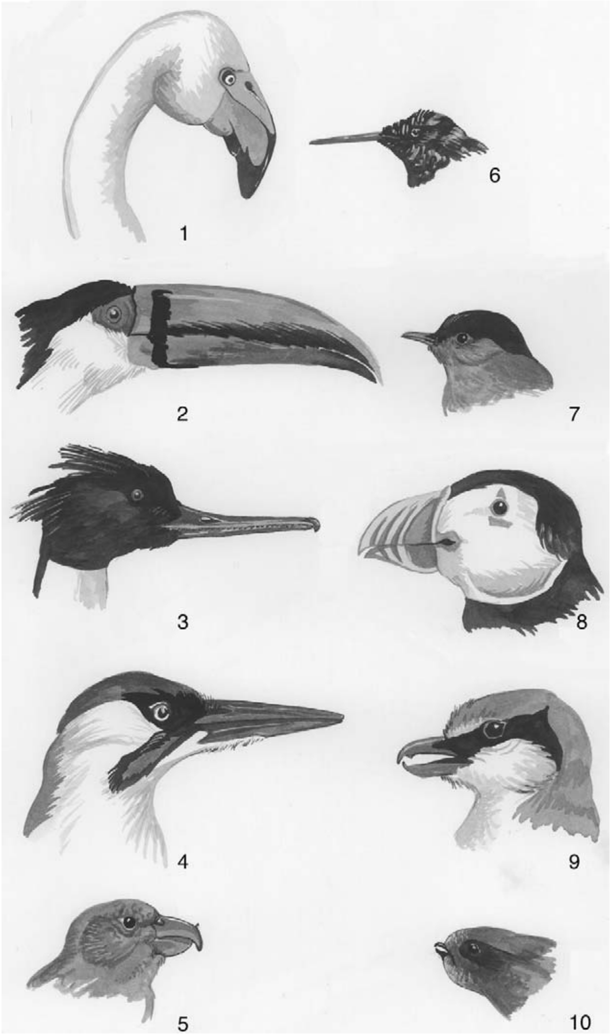


FIGURE 7 Examples of beaks used for various ways of feeding: 1, flamingo, modified for filter feeding; 2, toucan, elongate for reaching suspended fruit and secondarily decorated for social signaling; 3, merganser, serrated for gripping fish; 4, woodpecker, chisel-shaped for probing wood; 5, crossbill, specialized for extracting seeds from cones; 6, hummingbird, elongate for reaching nectar; 7, warbler, fine for picking up insects; 8, puffin, deepened for carrying several fish at once and secondarily decorated for social signaling; 9, shrike, hooked for tearing prey; 10, swift, wide for catching insects in flight (illustration by Su Gough).

borings with her long, pliant bill and the male chiseling the wood open with his shorter, stouter bill.

D. Migration

The most dramatic manifestation of birds' use of varying resources is migration. Huge areas of North America, Europe, and Asia become inhospitable to life in the winter, primarily because temperatures are low, the ground is snow covered, and waters are frozen; endothermic animals that could perhaps survive the physical conditions are often unable to survive in these regions because their food supplies are inadequate. Therefore, there are relatively few permanently resident birds. In the summer, however, there is an abundance of food and this presents an opportunity that is taken by hundreds of millions of birds from south temperate, subtropical, and tropical regions that move north to breed. Approximately 87% of the species breeding at 80°N in eastern North America are summer migrants, the proportion declining to 12% at 25°N. The same north–south pattern is seen in Europe but the proportion is consistently about 17% lower at equivalent latitudes because the warmer winters of western Europe allow more species to be year-round residents than in eastern America.

Not all movements are north–south. The comparatively warm (and damp) winter conditions on the western edge of Europe allow many waterfowl and shorebirds to winter there (such that British wetland sites tend to be more species rich in winter than in summer); these birds not only move north in summer but also move east, into continental regions where winter condition are intolerable. Summer visitors are generally more common in eastern than western Europe. Other species show altitudinal migrations, moving up into mountains to breed during short seasons of relative plenty. Within the tropics, savanna birds often have migrations that track the seasonally moving rainfall belts.

The seasonal variations in bird communities in particular areas may be complex. In the more benign parts of temperate regions, for example, such as the southern United States and Mediterranean Europe, winter visitors from the north and summer visitors from the south effectively replace each other. Thus, 52% of the wintering species at 25°N in eastern North America are visitors that have bred in the north; the proportion declines further north, to zero at 70°N. A similar pattern is seen in Europe, but there are about 10% fewer northern migrants at equivalent latitudes.

Summer migrants to North America mostly come from Mexico, Central America, the Caribbean, and northern South America rather than beyond the equator. These places are, if not forested, well wooded. Palearctic migrants come from similar latitudes, with 35% wintering wholly north of the equator and only 3% wholly south of it. In contrast to America, African migrants come almost entirely from savannas; these are not only much more extensive than forests in tropical Africa but also much more seasonal.

In much of Europe, summer visitors seem to occur particularly in grasslands and scrublands. In British and Irish woodlands, migrant passerines are most common in early successional stages and also in the upland oakwoods of Wales, where residents are less abundant than in lowland woods; migrant species are generally less abundant and less widespread than residents and use a less diverse array of nest sites. These patterns fit with residents, where they can survive the winter, preempting habitats, with migrants utilizing the summer surplus that the residents cannot fully take up.

There is much less land in the Southern Hemisphere than in the north. As a result, in both Africa and South America, although there is some migration between tropical and south temperate regions, it is less dramatic than the northern migration. As in the north, relatively few of the migrants cross the equator.

Although generalizations about migration can be made, there is much variation between species in details, as is apparent even from broad considerations of patterns of migration (Table II). Most migration is within-tropical, within-temperate, or tropical-temperate but there are a few species, such as American golden plover *Pluvialis dominica,* that move between temperate zones in the two hemispheres. At the extreme, some arctic terns *Sterna paradisaea* breed in the Arctic but winter in the Antarctic.

TABLE II

Percentages of Birds (Excluding Seabirds) Breeding in the Western Palearctic According to the Way in which Their Range Changes in Winter[a]

	%		
	Land	Freshwater	Coastal
Range is unchanged	44	25	8
North edge of range moves south (range contracts)	4	8	0
South edge of range moves south (so range expands)	4	5	0
Whole range moves south but summer and winter ranges overlap	18	46	46
Whole range moves south, summer and winter ranges do not overlap	30	16	46
Total number of species	399	63	26

[a] From Newton and Dale (1997).

There is variation even within species: Populations (and even individuals within populations) may differ in whether they are resident or migratory and, if migratory, how far they migrate and by what routes. "Leapfrog" migration is not uncommon. In the fox sparrow *Passerella iliaca,* for example, populations of Vancouver Island and the nearby mainland are resident; birds wintering in Oregon breed in the far southeast of Alaska and northern British Columbia; and those wintering in California breed further north and west, along the Alaskan peninsula.

E. Flightless Birds

A few bird species are flightless. Given that it is the demands of flight that are responsible for the restricted morphological diversity of birds, it is not surprising that some of these flightless species show marked departures from the typical avian body plan.

Many species that have colonized remote oceanic islands have become flightless. Such islands hold few, if any, other terrestrial vertebrates so there is a variety of vacant ecological niches available for birds to exploit, including those based on resources that, being neither patchy nor unpredictable, do not demand the power of flight for their exploitation. Nor is flight needed as a means of escape in the absence of predators. Among the flightless species on remote islands, rails (Rallidae) feature prominently. Rails generally lead a sedentary existence, using their powers of flight mainly for long-distance migration and to find isolated habitat patches. Island rails have typically achieved flightlessness, as have some other species (ratites in particular), by neoteny, i.e., the retention of juvenile features into adulthood. The key retained features are that the keel on the sternum (on which the massive flight muscles are inserted in flying birds) is absent (or much reduced), the legs are comparatively large, and the plumage is often looser than in typical birds, because the barbs that make up the vane of the feather lack the barbules that, in typical feathers, link adjoining barbs together to make the vane a continuous plate.

Other highly modified island forms, now extinct, were the dodo (*Raphus cucullatus*) of Mauritius and the two species of solitaire (*Pezophaps*) on Rodriguez and Réunion. Turkey-sized birds descended from doves, these species are typically pictured as exceptionally ponderous, although it has been argued that they displayed a marked annual cycle of fattening and that for much of the year they were far slimmer than usually portrayed. They had massive legs and bills and apparently occupied the niche of generalist herbivores, taking large seeds and foliage.

Apart from bats, New Zealand remained mammal free from its time of isolation from Australia 75–80 million years ago until colonized by man about 1000 years ago. In its forested landscape, moa (Dinornithidae) evolved: These were moderate to large birds, somewhat ostrich-like but much more sturdily built, and they usually held their necks horizontal rather than upright; they roamed scrubland and forests in search of seeds, fruit, and shoots. There were about 11 species, all extinguished by man. Kiwis (*Apteryx*), however, survive in small numbers. They are behaviorally and anatomically highly modified for feeding on soil invertebrates, with the nostrils on the tip of the elongate bill (unique among known birds). They are nocturnal. Apart from their locomotion being bipedal, they look remarkably mammal-like when rooting along the forest floor or across pastures in search of food. Another remarkable New Zealand species is the kakapo (*Strigops habroptilus*), an herbivorous flightless parrot; although a few other parrots are largely terrestrial, this is the only flightless species.

The moa-like elephant birds (Aepyornithidae) of Madagascar also evolved in isolation, with their final extinction apparently coinciding with the arrival of man. What of the surviving large flightless birds—the ostrich (*Struthio camellus*) of Africa, the two rheas (Rheidae) of South America, the emu *Dromaius novaehollandia* of Australia, and the three cassowaries (*Casuarius*) of Australia, New Guinea, and nearby islands? It is possible that all except the ostrich evolved in the absence of predators, although they have encountered them since; ostriches and rheas, living on open plains, can run at great speed. All have continued the bird habit of feeding on scattered food (fruit, seeds, flowers, shoots, and the more nutritious leaves, plus small animals in the case of emu and rheas), covering large foraging areas on their long legs and using the reach of their long necks to pick out food items.

The ecological gap left by the extinction of carnivorous bipedal dinosaurs led to the evolution of various giant flightless predatory birds but these seem to have been unable to survive competition from modern carnivorous mammals.

Waterbirds use two different techniques to swim under water, both of which result in a predisposition to the evolution of flightlessness. Some use their feet to drive themselves through the water. Thus, in addition to retaining the massive wing muscles required for flight, these birds must also have particularly powerful hindlimbs. More important, efficient movement through the water requires an elongate body, conflicting with the benefit for flight of a short, compact body. The powers of flight are therefore reduced in many foot-swimmers; species in isolated situations have become flightless, such as the Galapagos cormorant (*Nannopterum harrisi*) and the Lake Titicaca grebe (*Centropelma micropterum*).

The second underwater swimming technique is to use the wings, as do the auks (Alcidae) of northern seas. Since water is denser than air, the wing feathers need to be short if they are not to bend too much. Indeed, not only can the whole wing be much smaller but also it must be; otherwise, the wing muscles would not be strong enough to move it. It is not surprising that the Great Auk (*Alca impennis*) and other large species now extinct were flightless. In the Southern Hemisphere, even greater specialization for swimming occurred in the penguins (Spheniscidae). Their flippers are mostly made of muscle and bone, with the feathers on them being as short as those on the rest of the body. To prevent the bird from being too buoyant, the downy layer is not as well developed as one might expect for animals living in cold water; instead, penguins may carry much subcutaneous fat. Their bones tend to be solid. The elongation of their bodies is so great that they have to stand upright when on land. It can clearly be seen how giving up flight has liberated them from the constraints of body form that apply to most birds.

F. The Taxonomic Diversity of Birds

Approximately 9700 species of birds are currently recognized, which is more than twice as many as there are species of mammals. The great number of avian species is remarkable given the restrictions on morphological diversity imposed by flight, but it is undoubtedly a consequence of flight. As already considered, the mobility that flight provides allows birds to occupy finely divided niches. It also allows them to colonize isolated islands and habitat patches, where the original colonists may evolve into new species. [Note that bats, similarly equipped with the powers of flight, resemble birds in terms of their abundance/size relationships (Fig. 4) and they are the second most species-rich order of mammals, with approximately 950 species].

Table III shows the orders of birds recognized in traditional classifications of birds and those determined on the basis of studies of DNA hybridization. Many of the differences are trivial: The four orders united into the Struthioniformes on the basis of the DNA evidence have always been recognized as closely related; the splits of several orders generally correspond with long-recog-

TABLE III

Traditionally Recognized Orders of Birds and Those Based on DNA Hybridization, with Numbers of Species in the Latter[a]

	Traditional	Sibley and Ahlquist	Species
Rheas	Rheiformes	Struthioniformes	10
Ostriches	Struthioniformes		
Cassowaries	Casuariiformes		
Kiwis and moa	Dinornithiformes		
Tinamous	Tinamiformes	Tinamiformes	47
Grouse, pheasants, guinea fowls, and quails	Galliformes	Galliformes	214
Guans and megapodes		Craciformes	69
Swan, geese, ducks, and screamers	Anseriformes	Anseriformes	161
Woodpeckers and barbets	Piciformes	Piciformes	355
Jacamars		Galbuliformes	51
Hornbills	Coraciiformes	Bucerotiformes	56
Rollers, motmots, knigfishers, and bee-eaters		Coraciiformes	152
Hoopoes		Upupiformes	10
Trogons	Trogoniformes	Trogoniformes	39
Mousebirds	Coliiformes	Coliiformes	6
Cuckoos and hoatzin	Cuculiformes	Cuculiformes	143
Parrots	Psittaciformes	Psittaciformes	358
Swifts	Apodiformes	Apodiformes	103
Hummingbirds		Trochiliformes	319
Turacos and plantain-eaters	Musophagiformes	Musophagiformes	23
Owls	Strigiformes	Strigiformes	291
Nightjars and oilbird	Caprimulgiformes		
Doves and dodos	Columbiformes	Columbiformes	313
Sunbittern, bustards, cranes, Kagu, rails, etc.	Gruiformes	Gruiformes	196
Button-quails		Turniciformes	17
Sandgrouse	Pteroclidiformes	Ciconiiformes	1027
Shorebirds, gulls, terns, and auks	Charadriiformes		
Grebes	Podicipediformes		
Cormorants, pelecans, gannets, etc.	Pelecaniiformes		
Hawks, eagles, falcons, secretary bird, etc.	Falconiformes		
Storks, herons, ibises, etc.	Ciconiiformes		
Petrels, shearwaters, and albatrosses	Procellariiformes		
Divers	Gaviiformes		
Penguins	Sphenisciformes		
Flamingos	Phoenicopteriformes		
Perching birds	Passeriformes	Passeriformes	5712

[a] Sibley and Ahlquist (1990).

nized subordinal divisions. The most fundamental difference is not just that 10 apparently diverse orders are united into the Ciconiiformes by the DNA evidence but also that some of these orders have not retained their integrity as major subdivisions within the combined order. Further work is needed before these changes can be accepted or rejected.

The classification based on DNA hybridization groups the orders of modern birds into two “infraclasses,” Struthioniformes plus Tinamiformes (Eoaves) and all the others (Neoaves). This is a long-recognized division based on the peculiar structure of the palate and other anatomical characteristics of the first group. Eoaves are often referred to as ratites, although strictly speaking this term should not include the tinamous since it refers to the condition of the sternum in the struthioniform birds, which is flat, without the keel on which the flight muscles of most birds (including tinamous) originate.

Both the traditional and the DNA-based classifications recognize more orders of birds than the 21 usually recognized for mammals. Given that birds are substantially less diverse than mammals, this is paradoxical. It reflects the fact that taxonomic categories at the same formal level in the hierarchy are not comparable in different groups and that ornithologists have adopted classifications that specialists in other disciplines would regard as inflated. (Entomologists would probably reduce most ornithological families to the rank of genus.)

The outstanding feature of avian taxonomic diversity is that 59% of the species belong to one order, the Passeriformes. In contrast, the most species-rich order of mammals (the rodents) comprises only 42% of the species. The passerines are "typical birds." They are characterized by a particular form of the foot, well suited to perching, but have no obvious single characteristic that would appear to open up ecological opportunities in the way that the special gnawing incisors seem to have done for the rodents. Perhaps their numerical dominance is simply related to their small size, given that most birds are small (see Section I, B). An additional peculiarity is that the two suborders of passerines are disparate in size, with 80% of the species falling into the suborder Passeri (or oscines), the true songbirds, and only 20% in the suborder Tyranni (or suboscines). Speciation in songbirds often seems to involve the evolution of song differences between formally conspecific populations. Perhaps the complex vocal apparatus of oscines has promoted high rates of speciation, leading to this suborder (and thus the order to which it belongs) being remarkably species rich.

Another peculiarity of birds, particularly if one considers that avian orders are comparable to taxa of lower rank in other animal groups, is that only a few of the orders contain very few species. The Coliiformes is the only order with less than 10 species, whereas six mammalian orders have fewer than 10. The most taxonomically isolated bird species, generally given families of their own, include the secretary bird *Sagittarius serpentarius,* a large ground-living raptor of African savannas; the hoatzin *Opisthocomus hoazin,* a leaf-eating oddity related to cuckoos which inhabits permanently flooded forests in South America, with an enormous specialized crop for processing its unusual diet; the kagu *Rhynochetos jubatus,* a strange, short-legged flightless gruiform endemic to the forests of New Caledonia; and the oilbird *Steatornis caripensis,* a Neotropical caprimulgid that feeds on fruit by night using sonar to navigate around the dark caves into which it retreats by day. Despite their peculiarities, none of these is as divergent from the rest of their class as is the aardvark *Orycteropus cafer,* placed in an order all its own among the mammals.

II. HABITAT AND BIRD SPECIES DIVERSITY

A. Differences between Habitats in Bird Species Diversity

It is a matter of common observation that habitats in the same region differ in the species diversity of the bird communities living in them. Habitats with higher diversity also tend to hold larger numbers of individual birds. However, there have been few systematic surveys of bird communities across all the habitats in a region and even fewer with the extent of replication used for the data in Fig. 8, in which each habitat was represented by at least 12 sites (with a mean of 72 sites/habitat). This confirms the following generalization: Woodlands, scrub, and carr contain large numbers of species, whereas open downland contains very few. Machair, flat plains based on deposits of calcareous sand fringing parts of the western seaboard of northern Scotland, is a conspicuous exception, having large numbers of birds despite being completely unwooded. This may be the result of the extreme small-scale heterogeneity of machair lands: There may be a transition from mobile sand dunes to stabilized grassland, from very dry to very wet conditions, and from undisturbed to cultivated ground within distances of tens of meters.

The distributions of the points on the two axes of Fig. 8 are correlated: Woodland tends to have many of the more ubiquitous species (occurring in more than 75% of the sites) and the less ubiquitous (occurring in only 25–75% of sites), whereas open downland holds few of either. The scatter around the correlation, however, reveals further differences between habitats. For example, scrub is an ephemeral habitat, typically occupied by species that are good colonists and which therefore tend to be found in the majority of sites; as a result, two-thirds of the species found in more than 25% of scrub sites are actually found in more than 75% of sites.

B. Bird Species Diversity in Similar Ecosystems on Different Continents

If habitat is a major determinant of bird species diversity, one would expect to find similar levels of diversity in similar vegetation and landscape types on different

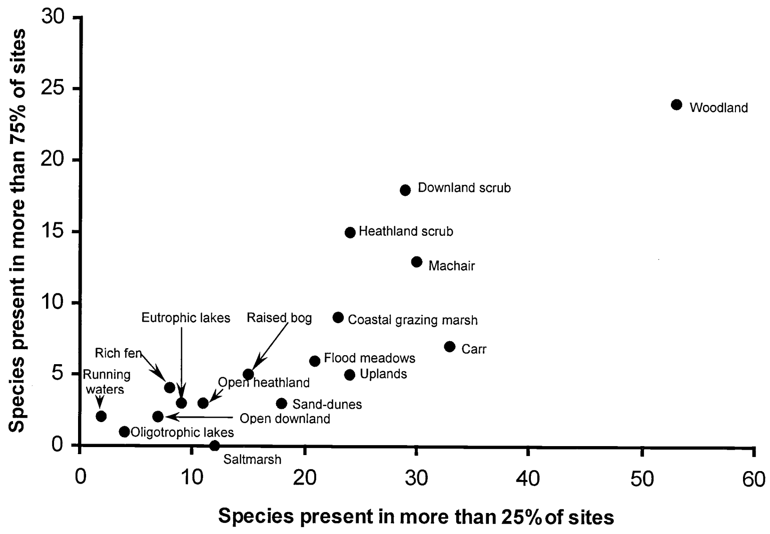


FIGURE 8 The numbers of species in British bird communities, according to how many are found on more than 25% of sample sites for each habitat and how many on more than 75% (from data in Fuller, 1982, with permission of the British Trust for Ornithology).

continents. Often one does, but sometimes one does not (Table IV). Even where total species numbers are similar, the numbers of species within individual feeding categories may differ. Thus, although numbers of bird species in most feeding categories are similar in Mediterranean scrub in Provence (France), Chile, and California, Provence has no nectarivores and only 4 granivores, in contrast with 3 and 14, respectively, in California and 2 and 6 in Chile. Again, the species richness of semiarid habitats is similar in North America and Australia, but the bird communities are clearly different in detail, partly perhaps as a result of different evolutionary histories but partly because rainfall is more erratic in Australia than in North America. Thus, the habitats are similar but the overall environmental conditions differ significantly. Despite differences in species number, the number of species in the various feeding categories tends to depend more on habitat itself than on what continent the habitat is in (Table V).

TABLE IV

Numbers of Bird Species in Similar Habitats on Different Continents[a]

Habitat	No. of species	
Peatlands	33 (Finland)	18 (Minnesota)
Desert	57 (Arizona)	61 (Argentina)
Shrub desert	5.5 (Australia)	6.3 (N. America)
Mediterranean scrub	30 (California)	28 (S. Africa)

[a] From a compilation by Schluter and Ricklefs (1993).

C. Bird Species Diversity and Habitat Complexity

Although it is not a universal rule, most studies have tended to confirm the discovery by R. H. and J. W.

TABLE V

Numbers of Bird Species Recorded from Mediterranean Scrub and Wet Tropical Forests in Africa and South America, Grouped into Foraging Categories[a]

	Forest		Scrub	
Foraging category	Africa	South America	Africa	South America
Arthropods from foliage	77	55	4	3
Seeds and fruit	20	49	3	1
Arthropods from sallies	21	21	1	1
Trunk, bark, etc.	13	31	0	1
Ground	13	21	7	10
Raptors and scavengers	15	13	2	4
Nectar	1	12	1	1
Aerial and crepuscular	5	3	1	1
Total	165	205	19	22

[a] Data collated by Schluter and Ricklefs, (1993).

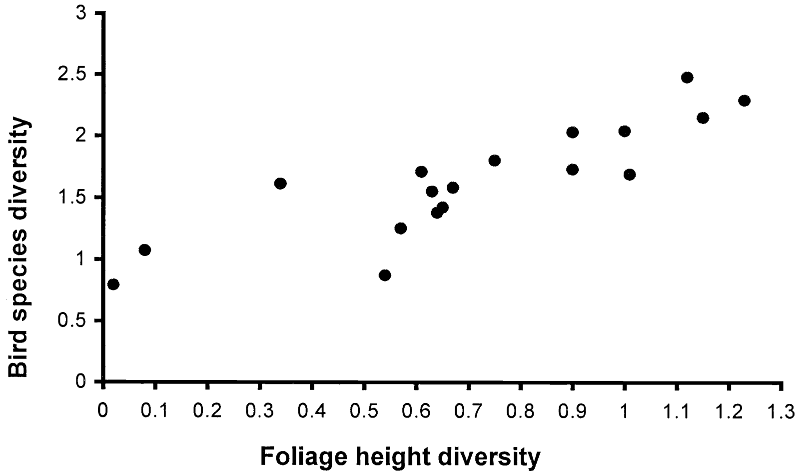


FIGURE 9 The relationship between songbird species diversity and foliage height diversity in Scottish woodlands. Some points are averages over 2 or 3 years. Both diversities are measured using the Shannon index. (redrawn from Moss, 1978).

MacArthur that bird species diversity is positively correlated with foliage height diversity, at least in scrub and woodland habitats (Fig. 9). That is, bird species diversity is greatest where there is an equal volume of foliage at all heights in the woodland. An early finding was that data points from Australia fell on the same line as those from the original study in North America. However, MacArthur discovered that data from Panamanian and Puerto Rican sites only fell on the same line if the measure of foliage height diversity was adjusted by changing the height subdivisions on which it was based. His interpretation was that Panamanian birds were effectively treating the forests as being divided into more layers than were North American birds, whereas Puerto Rican birds were treating them as being more coarsely divided. Similar variations have been found elsewhere and may be related to the productivity of the different areas. Where there are more resources available, species may divide them into narrower ecological niches; perhaps more simply, more resources allow more individual birds to live in an area and thus allow even relatively scarce species to maintain viable populations.

Along an altitudinal gradient from lowland rain forest at 400 m to elfin woodland at 3600 m in the Peruvian Andes, bird species diversity and foliage height diversity decline approximately in parallel. However, the magnitude of the decline differed according to the diets of the birds in question. It was more than fivefold for insectivores but only two- or threefold for frugivores, whereas the species diversity of nectarivores was similar at all altitudes. These differences may be explained in terms of the resources available to different groups: Whole groups of insects are absent at higher altitudes; less fruit is produced at higher altitudes but a higher proportion is available to the birds because there are fewer mammalian competitors than there are at lower altitudes; and because of the distribution of flowers on the vegetation, the availability of nectar only increases slightly with foliage height diversity and, although the annual production of nectar is greater at lower altitudes, it is highly seasonal and therefore can support no more birds than the less abundant but more constantly available nectar of the high-altitude forests. Thus, foliage height diversity is a crude measure of the diversity of ecological niches available to birds but it may become less satisfactory when one considers individual trophic guilds.

Since foliage height diversity is only a rough measure of niche diversity, it is easy to find other features of the habitat that influence bird species diversity. Thus, a survey of various sites in the Americas showed that although there was a good correlation between bird species diversity and foliage height diversity, the percentage vegetation cover of the ground also influenced the number of bird species. In other words, the total volume of vegetation was important, not just how it was structured. In hardwood forests of New Hampshire, for example, tree species differ in their value to birds: Some hold more food than others and some are easier to search (e.g., yellow birch *Betula alagheniensis,* with small leaves and short petioles); white ash *Fraxinus americanus* has a distinct foliage structure that allows fewer foraging methods to be used than do other species. Furthermore, floristic

FIGURE 10 The relationship between bird species diversity and the percentage of the basal area of the forest comprising standing dead wood in Swedish woodlands (reproduced with permission from Nilsson, 1979).

diversity is important: Even a few conifers in a hardwood forest allow in birds that specialize on such trees. In south Swedish forests the amount of standing dead timber (which is largely determined by how the forest is managed) chiefly determines bird species richness (Fig. 10), with the density of low trees and shrubs having only a slight effect.

The most extreme case of the influence of factors other than foliage height diversity on bird species diversity occurs in forests in Patagonia, in which the relationship between birds and foliage diversity appears to be reversed (Fig. 11). The reason for this is that the forests are of two quite different types and the type with the greater foliage height diversity (*Nothofagus* forest) has relatively few birds because it comprises patches that are isolated from other such forests by 2000 km of steppes and forests; furthermore, these forests are at high elevations, with harsh climates. The abundance of bamboo, which appears to support few insects, in the understory of some of the *Nothofagus* forests may also be important. Note that even these species-poor forests hold more species of birds than grasslands in the same region (Fig. 11), fitting the usual pattern.

Horizontal complexity is also important, with more varied places holding more bird species. Edges between habitats may be bird rich for this reason, although "edge effects" are variable both in nature and in cause. In Britain, woodland birds are generally more abundant and more diverse at edges, probably because foliage height diversity is greater there than deep in the woodlands, where the lower levels of the vegetation tend to be shaded out. On Finnish islands, bird species diversity is lower on the edges of woodlands than deeper within perhaps because of the importance of physical protection during the early nesting period, when most of the trees have not yet begun to leaf.

D. Foraging Niches and Habitat Structure

Classic studies of the foraging behavior of tits in broadleaved woodlands in southern England were conducted by David Lack and colleagues. In summer, all species tend to feed largely on the glut of caterpillars in the

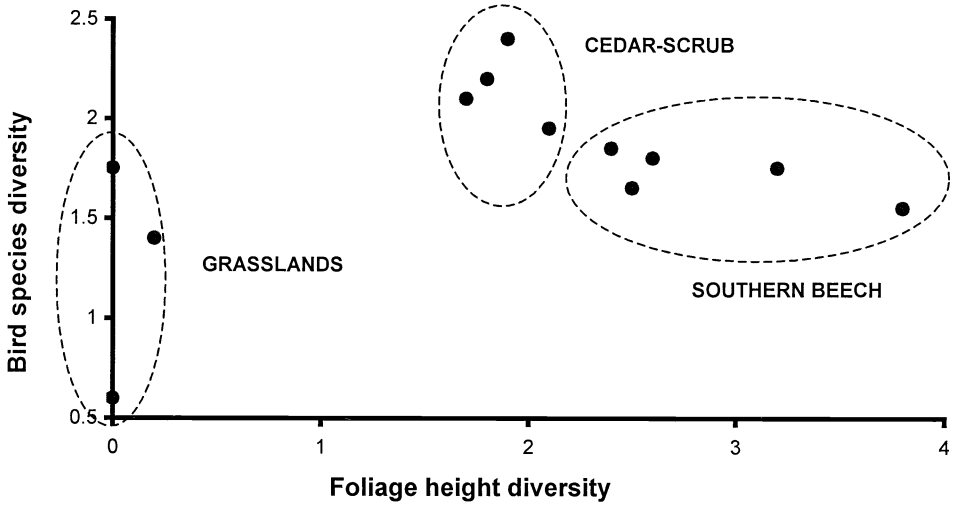


FIGURE 11 The relationship between bird species diversity and foliage height diversity in three contrasting habits in northern Patagonia (reproduced with permission from Ralph, 1985).

canopy; in winters during which the mast of beech *Fagus sylvatica* is abundant, they descend to the ground to feed on it. However, during the average winter their feeding niches are distinct. The great tit *Parus major* feeds mainly on the ground on beech mast, other large seeds, and larger insects. The blue tit *Parus caeruleus,* only about half the weight of the great tit and thus much more agile, frequently hanging upside down to reach its food, feeds high up in the trees, searching for small insects among the twigs, buds, and leaves. The less common marsh tit *Parus palustris,* although no heavier than the blue tit, has a larger beak and feeds on middle-sized insects, which it gathers in the shrub layer or from the twigs and branches of trees, lower down than where the blue tit feeds. Similar studies of other birds in many parts of the world have given similar results: Related species coexisting in the same forests and using broadly similar foods tend to differ in respect of the parts of the trees in which they forage. However, foraging height is not the only factor. Thus, the coal tit *Parus ater* in English broad-leaved woodland tends to forage at approximately the same height as the marsh tit, but it is a smaller species with a finer bill and it feeds on tiny insects that it finds in crevices on the bark. In addition, it is scarce in broad-leaved woods because it is a conifer specialist; its numbers are significantly enhanced when conifers are present—another example of the importance of floristic diversity for bird species diversity.

It is such differences in foraging site and food taken that allow different species to coexist in the same habitat. Introduced birds on Hawaii provide an unplanned experimental demonstration of how differences allow coexistence: Some species persisted for more than 10 years but then became extinct; extinction was particularly likely for species that had morphologically similar relatives (presumably taking similar diets) living in the same forest. Foraging niches sometimes, but not always, shift in apparent response to the presence or absence of competitors: Willow tits in northern Finland avoid the foraging positions of great tits *P. major* and crested tits *Parus cristatus* when in mixed flocks, but if either of these is absent from a flock the willow tits use the vacated niche.

Different strata in forests differ not just in which species occupy them but also in their overall species richness. In the example shown in Fig. 12, the forest floor, relatively uniform and simple in structure, supports few species—as does the high canopy, which is also relatively simple in structure. The middle layers support many more species because they are more complex, presenting opportunities for foraging on leaves, branches, bark, dead wood, epiphytes, termite nests, and accumulations of dead leaves in tree forks and for sallying from branches to capture insects flying in the open spaces below the canopy. Underlying the overall pattern of species richness are marked differences between feeding classes (Table VI): Mast is found either high in the canopy or fallen to the forest floor, so mast feeders occur only in these layers; gleaners are also

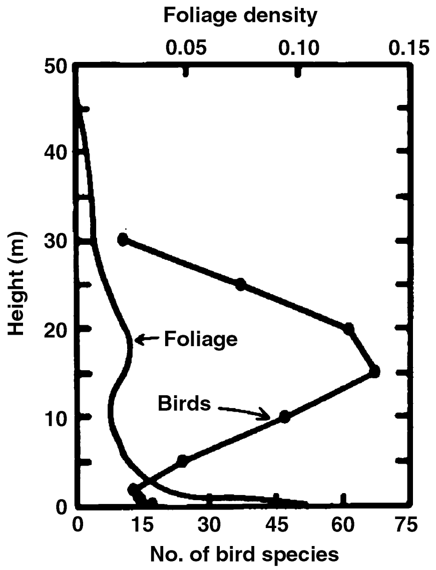


FIGURE 12 The relationship between bird species richness, foliage density, and height above ground within a Peruvian forest (reproduced with permission from Terborgh, 1980).

TABLE VI

Number of Species of Different Feeding Guilds Found at Various Heights in Peruvian Forests[a]

	Height class (m)					
Guild	Ground	0–5	5–10	10–15	15–20	>20
Mast	3					3
Fruit	3	2	2	5	3	7
Nectar		6	2		2	1
Insects						
Glean	5	10	5	2	7	2
Sally	1	2	5	7	3	4
Bark			1	7	4	1
Fruit/predator				1	3	1
Fruit/insects				5	9	2
Fruit/insects/nectar					4	4

[a] From Terborgh (1980).

mostly found low down or high up, working over the continuous substrates of the forest floor or the canopy; and salliers tend to occupy middle levels, where the open structure allows them to see and pursue flying insects. The canopy is a much patchier habitat than the understory because the crowns of individual trees occupy large areas, whereas the understory is composed of much larger numbers of plants; thus, the canopy presents particular opportunities to adaptable generalists that can switch, for example, from feeding on fruit to feeding on insects depending on what is immediately available.

E. Bird Species Diversity and Succession

Where sheep grazing is light enough to allow it, scrub invades open hill land in Wales, eventually developing into mature sessile oak (*Quercus petraea*) woodland. This vegetational succession is accompanied by an almost monotonic increase in the numbers of both individual birds and bird species (Fig. 13). Most of the change involves increased vertical structure, although the latest stages result from gaps developing in the woodland canopy, with small patches of open ground and scrub. As chalk downland in southern England is succeeded by ever-denser scrub, species numbers increase monotonically from 1 or 2 per study site to approximately 15 (Fuller, 1982). More generally in Britain, the scrub that develops on downland and heaths and the carr that develops on rich fens hold more species overall and more of these species are found on a high proportion of individual sites (Fig. 8).

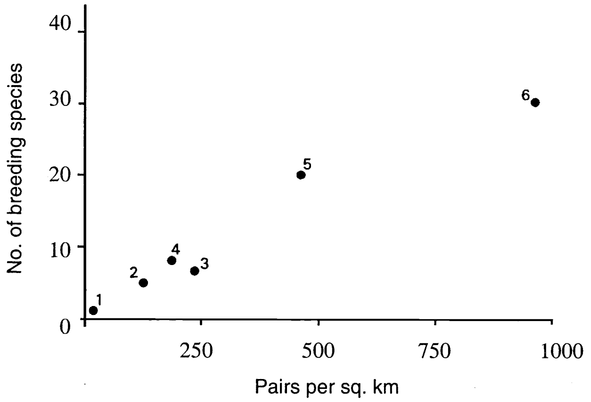


FIGURE 13 The relationship between the number of breeding bird species and the total population density in sessile oak woods in Wales in various stages in the succession: 1, open ground; 2, low scrub; 3, medium scrub; 4, high scrub; 5, mature woodland; 6, mature woodland with gaps and scrub (reproduced with permission from Jones, 1972).

The increase in bird species diversity may slow in later stages of succession. In a subalpine northern American sere comprising meadow, aspen (*Populus tremuloides*), fir (*Abies lasiocarpa*), and spruce (*Picea engelmanni*), the number of bird species averaged over 3 years was 3, 17, 22, and 21, respectively. Indeed, although as yew *Taxus baccata* woodlands in southern England mature they gain species steadily (10 species at 60–80 years old and 20 species at 100–200 years), the succession from open scrub with small stands of trees to the young closed yew woodland is marked by a decline from 26 to 10 bird species. Polish deciduous forest stands developing after clear-cutting reach their maximum bird species diversity when 100–150 years old but do not do so monotonically: Following a peak at 15–20 years, diversity declines for the next 15–20 years. Local factors may be important in determining the relationship between bird species diversity and habitat gradients, such as the gradient from very low matorral (0.5 m high) to mature forest of holm oak (*Quercus ilex*): Whereas the number of bird species increases along the vegetation gradient in mainland France, it decreases in Corsica (Table VII) because Corsican forests are very restricted and the island simply lacks many forest species found on the mainland, where they occur in large blocks of continuous forest. An additional complication to the story of Mediterranean holm oak forests is that, although the number of bird species in an area is reduced by fire, the reduction is less than expected from the change in vegetation structure. Some characteristic forest birds persist in the open areas and matorral created by the fire either because individual birds simply occupy the same territories or because they spill over from populations in nearby unburned forest. Human management can produce habitats in which the normal pattern is disrupted. For example, when English woodlands are coppiced (the trees are cut to the ground but allowed to regenerate from the stumps), the dense, bushy regrowth is colonized by numerous species, but

TABLE VII

Numbers of Bird Species Breeding in Corsica and in Comparable Areas in Mainland France in Relation to the Stage of Succession, from 1 (Very Low Matorral, 0.5 m High) to 6 (Mature Forest of Holm Oak *Quercus ilex*, 2.5 m High)[a]

Stage	1	2	3	4	5	6
No. of species						
Corsica	29	24	32	24	23	18
France	11	15	11	19	26	23

[a] From Blondel (1991).

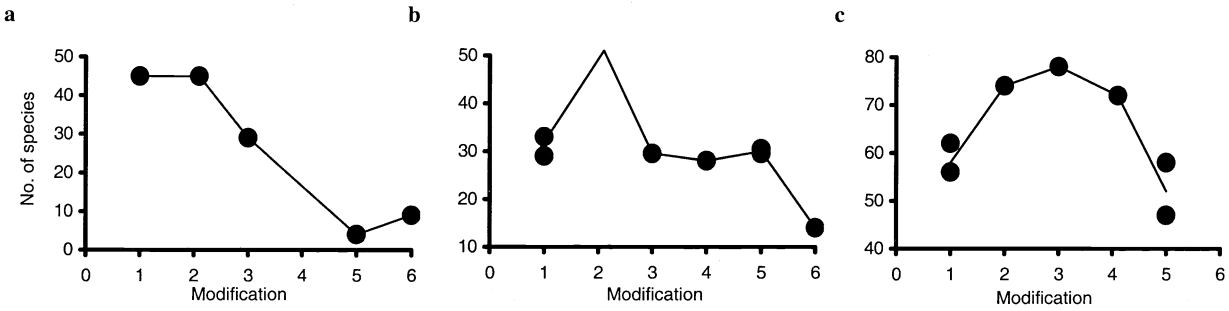


FIGURE 14 Species richness of (a) birds, (b) butterflies, and (c) leaf-litter ants along a gradient of increasing habitat modification in Cameroon. 1, Near-primary forest; 2, old secondary forest; 3, partial manual clearance, plus plantation; 4, partial mechanical clearance, plus young plantation; 5, complete clearance, with young plantation; 6, manually cleared farm fallow (redrawn from Lawton *et al.*, 1998).

many of these are lost as the coppice ages and the tall regrowth forms a dense canopy that shades out the majority of the lower layers, producing a uniform, structurally simple habitat.

Even where the relationship between bird species diversity and habitat follows the typical pattern, this may not be simply because structurally less complex habitats provider fewer niches. Thus, bird species richness in cleared areas in Cameroon is much less than that in the forest (Fig. 14), but this may be largely because the sources of open-ground species are savannas and grasslands that lie well beyond the surrounding forest so that the cleared areas may not have been fully colonized. (Note that this case also illustrates that just as different trophic groups of birds differ in their response to vegetation structure so do different taxa of animals generally: Butterfly species richness is affected by forest clearance in much the same way as that of birds but leaf-litter ants are more species rich in partially cleared and second-growth forests.)

In contrast to the absolute level of species diversities, the turnover of species during succession has been little studied. In Polish oak-hornbeam forests, there is a rapid shift in the early stages from species characteristic of open ground to those characteristic of scrub, but the rate at which species drop out of or enter the community slows as the succession progresses. In Finnish coniferous forests, in contrast, turnover peaks not at the start of succession but at 5–25 years.

F. Bird Species Diversity and Productivity

Many of the habitat differences that seem to influence bird species diversity are probably associated with differences in total biological productivity. Even with habitats of uniform physical complexity, such as the lakes of a region, differences in productivity are associated with large differences in the numbers of both individual birds and species present.

III. GEOGRAPHICAL VARIATION IN BIRD SPECIES DIVERSITY

A. The Relationship between Bird Species Diversity and Habitat Diversity

Different habitats have different bird communities. The greatest differences within geographical regions are between marine and terrestrial habitats since life on land and at sea requires different specializations. On land, the greatest differences are between bird communities of wooded areas and those of open country, just as these broad habitat categories show marked differences in overall species richness. However, the differences go beyond species richness: Grasslands differ from woodlands not just in that most of the woodland species are absent but also in that they have their own grassland specialists. The suites of species occurring in different habitats may be similar in size (number of species) but quite different in composition. For example, the numbers of nonmarine species occurring in more than 25% of sites in three British coastal habitats are 18 in sand dunes, 15 on rocky coasts, and 12 on salt marshes, but only 4 species are common to all three habitats and 6 common to sand dunes and rocky shores, whereas there are 7, 10, and 2 unique to sand dunes, rocky shores, and salt marshes, respectively.

Related species often differ in the habitats they occupy. For example, in Britain great tits *P. major* live in both deciduous and open conifer woods, blue tits *P.*

caeruleus and marsh tits *P. palustris* only in deciduous woods, and coal tits *P. ater* almost only in conifer woods. In central Sweden, marsh tits again occupy deciduous woods and willow tits *P. montanus* live in conifers but on the nearby Åland Islands, from which marsh tits are absent, willow tits occupy both.

Given such differences, one would expect bird species diversity to be correlated with habitat diversity in an area. Although this is undoubtedly the case, it has rarely been documented. However, a study of the boreal birds occurring on isolated mountaintops and mountain chains in western North America showed that 91% of the variation in species number could be accounted for by differences in habitat diversity.

B. Bird Species Diversity and Altitude

The decline in bird species diversity with increasing altitude in the Peruvian Andes (see Section II,C) is typical not just of forested regions but also of open habitats and not just of the tropics but also of temperate and boreal regions. Everywhere, there are fewer species of birds at higher altitudes—unless the higher levels have escaped habitat destruction wrought by man in the lowlands or unless the contrast is between different habitats, such as lowland savanna and montane forest. One cause of such gradients is that the structural complexity and the floristic diversity of the habitat are reduced at higher altitudes. This is true of both forests and open habitats; indeed, forest gives way to open habitats at sufficiently high elevation.

The cooler, windier, and wetter conditions at high altitudes are no doubt responsible for the less luxuriant vegetation and probably also affect the birds directly so that only species capable of withstanding them can survive. Some mountains may also be so isolated from others that they resemble islands, lacking some species because they have never reached these mountains or because they have not recolonized following a chance extinction (see Section V).

C. Bird Species Diversity and Latitude

As with so many other groups, there are more bird species in the tropics than toward the poles (Fig. 15), although some groups do not conform to this generalization: Only 3 of the 17 species of penguins (Spheniscidae) breed within the tropics (where marine productivity is low) and sandpipers (Calidridinae) breed mainly in the boreal tundra and taiga (utilizing the summer flush of productivity, especially in wetter areas). The trend occurs in all habitats but may be steeper in forests than in savannas and prairies.

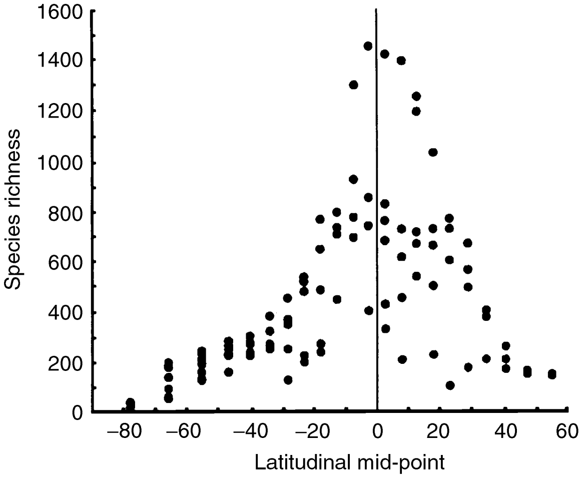


FIGURE 15 The relationship in the New World between the species richness of birds breeding in a square and the latitude of the square (redrawn from Blackburn and Gaston, 1996b).

The broad trend is modified by local contingencies. In Argentina, the trend is steepest where forest gives way to savanna. In Africa, the reduction in species number as one moves away from the equator is steeper to the north than to the south, in accord with gradients of aridity and vegetation diversity. Northern finnish peat bogs hold more breeding bird species (and individuals) than southern ones, probably because they have willows (*Salix*) and therefore are more heterogeneous than the willow-free southern bogs. In Sweden, breeding passerine species are scarcer in northern than in southern bogs but waders are more diverse, with the northern bogs having more food-rich pools. In heavily urbanized locations in Finland, there is no latitudinal gradient in species diversity in winter because the birds are wholly dependent on man for food in such places. Mountain regions tend to be species rich because of their topographical and habitat diversity. Coastal areas differ from inland areas in diversity of nonmarine birds for reasons that are not clear: In North America they are comparatively poor in terms of breeding species, but in parts of Europe they are richer than inland squares.

Even at the continental scale, there are usually major departures from simple north–south trends. The pattern of aridity (and associated vegetation) largely determines bird species diversity in both Australia and Africa. In North America (Fig. 16), the east is richer in breeding bird species than the west, although the far west is

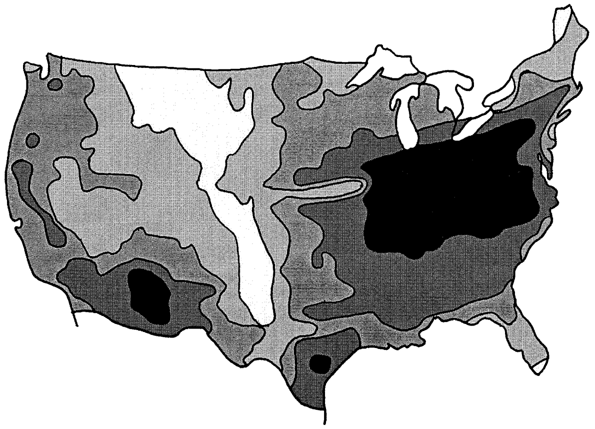


FIGURE 16 The distribution of species richness of insectivorous birds breeding in the United States. The depth of shading represents richness, from less than 70 species (pale) to more than 100 (dark) (redrawn with permission from Schall and Pianka, Geographical trends in numbers of species. *Science* **201**, © 1978 American Association for the Advancement of Science.).

richer than the Midwest, especially in the mountains and the southwestern desert; in the east, there is a midlatitude peak in richness. The latter is also seen in Europe. It is not simply a gross pattern that might be produced by differences in the proportion of richer and poorer habitats because it is observed within individual habitats.

Migration has major impacts on bird species diversity in many regions and habitats: It is substantially lower in the Arctic and much of the north temperate zone in winter than in summer and must vary considerably in African savannas because of migration both between Africa and the Palearctic and within Africa. This can modify geographical patterns of species richness. Thus, the difference between African forest and savanna is diminished, although by no means eliminated, in the northern winter because almost all the migrants move into the savanna; however, within the savanna the geographical pattern is maintained because areas rich in resident species are also areas to which most migrants move. In both Europe and eastern North America, the midlatitude peak of species richness in summer is eliminated in winter, when there is a simple north–south gradient.

D. Comparisons with Other Animals

There have been few direct comparisons between the geographical patterns of species richness of birds and those of other animals. In North America, a midlatitude peak in richness is also seen in mammals and amphibians, although the peak is further south, especially for amphibians. Reptiles show a simple north–south gradient. The decline in species richness in the north is steeper in mammals than in breeding birds, even steeper in amphibia, and steepest in reptiles. (Note, however, that many birds move out of the north in winter.)

Over North America as a whole, the species richnesses of birds and mammals are positively correlated but birds are not correlated with lizards and are negatively correlated with turtles and snakes. In Australia, there is a similar divergence: Birds are positively correlated with marsupials but negatively with lizards (which are species rich in hot, dry conditions, in which only specialist birds live).

Thus, although birds resemble other animals in showing a general tropical–polar gradient in species richness, local factors may affect them differently.

E. Causes of the Latitudinal Gradients in Species Diversity

Avian data have been widely used to illuminate possible causes of the tropical–polar gradient in species diversity. It is possible that it depends in part on there being more niches available in the tropics. Tropical vegetation is floristically more diverse and, in forests at least, structurally more complex. In North America, the midlatitude peak in bird species diversity is approximately matched by tree species diversity, as are the broad east–west trends. However, the relationship between the species diversity of birds and that of trees in North America is asymptotic; above a certain level, the richness of tree species appears not to influence that of birds. In Amazonian forests, 15% of bird species depend on habitats arising through dramatic fluctuations in river levels which do not occur in temperate regions. Furthermore, 34% of Amazonian birds belong to foraging guilds that are not found in temperate regions, such as obligate dead-leaf gleaners, obligate ant followers, and year-round frugivores; the rain forest's richness and comparative constancy provide opportunities lacking in temperate forests. However, this is not the whole story because even where guilds are found in both Amazonian and temperate forests they contain fewer species in the latter.

The North American midlatitude peak in breeding bird species diversity has been interpreted in terms of migration. At low latitudes, there are many year-round residents but at midlatitudes there are fewer; therefore, the summer provides many opportunities for migrants to pour in and to exploit the seasonal abundance of productivity. There is indeed evidence that insect food, although seasonally much more restricted in midlatitudes than in the south, is much more abundant when it does occur. Furthermore, it is migrants that feed on the insects that are active in summer that contribute most to the midlatitude peak; resident species such as woodpeckers, which have less seasonally variable food supplies, are not more numerous in midlatitudes than in the south. The hypothesis explains the decline in species richness in the far north (where there is also a summer flush of resources) in terms of constraints imposed by the shortness of the summer and the long journeys that migrants need to make to get there.

Another possibility is that the tropics are species rich because they include greater land areas, allowing more opportunities for speciation and reducing the likelihood of extinction. Although there is no correlation between the species richness of New World birds and land area at a given latitude outside the tropics, there is such a correlation if one excludes those species that are primarily tropical. It appears, therefore, that there is an underlying relationship between species richness and land area in the New World but that the very high richness of the tropics superimposes a wider gradient because some tropical species extend into subtropical and temperate regions.

An alternative hypothesis is that species richness is correlated with available energy in an area, either directly (warmer places allow animals to function more effectively) or indirectly, via productivity (higher productivity supports more individuals and therefore more species). Avian species richness in North America is correlated with potential evapotranspiration, which is a measure of available energy, but less well with actual evapotranspiration, which is more a measure of primary productivity. This supports the direct energy hypothesis. In Britain, temperature gradients are north–south in summer but, because of the influence of the North Atlantic Drift, east–west in winter. Data on the distributions of resident and wintering birds have been used to test the species energy hypothesis, predicting different patterns of species richness in summer and in winter. Early analyses provided support for the direct effect of energy, including indications that correlations with temperature were greater for smaller birds (which are more susceptible to low temperatures). Recent work has not confirmed this result. Nor has it found seasonal changes in distribution that match the predictions of the direct energy hypothesis. Summer temperature is the best predictor (of various climatic and topographical variables) of bird species diversity in Britain, in both summer and winter. This suggests that birds in winter depend largely on productivity stored from the summer (in the form of seeds, nuts, and invertebrates).

IV. SPECIES REPLACEMENTS THROUGHOUT THE WORLD

A. The Ranges of Birds

The barn owl *Tyto alba* is found in every continent of the world except Antarctica, and it is widespread in most of them. Most birds are much more restricted in terms of habitats, climate zones, and continents. Some occur on single small islands or isolated mountaintops. Differences in which species occur in which places are just as much a part of biodiversity as is the species diversity of individual places.

B. Species Replacements

Two species may have different geographical ranges because they live in different habitats that are found in different parts of the world. Alternatively, they may occupy similar habitats but simply have evolved in different places. Thirdly, they may originate in isolation from each other, subsequently come together, but be ecologically so similar that competition prevents them from coexisting, except perhaps in limited areas (Fig. 17). If they differ sufficiently in feeding habits or habitat, coexistence is possible but the extent of this varies between different sorts of birds, between breeding and winter quarters, and (within a taxon) between continents (Table VIII). The occurrence of closely related species pairs that overlap only partially, if at all, such as illustrated in Fig. 17 is common in Europe and is probably the result of the range of an original species being split in two by the great climatic fluctuations of the Pliocene and Pleistocene. The Great Plains of North America are another zone in which such pairs of species meet.

C. Centers of Endemism

The ranges of 27% of the species of land birds in the world extend less than 50,000 km^2. Such restricted-range species tend to occur in particular areas: Some 221 “endemic bird areas” (EBAs), each with two or more restricted-range species, have been identified. The concentration is such that 2% of the earth’s land surface contains 20% of the bird species. The tropics contain 76% of the EBAs but this may be no more than a reflection of the generally high species diversity in the tropics. Reflecting the importance of isolation in speciation, almost half of the EBAs are on islands, half of which are small and oceanic. Continental EBAs are often in mountainous regions, where both mountains and valleys can form barriers to the dispersal of species that

FIGURE 17 Ranges of nightingale *Luscinia megarhynchos* (dark shade) and sprosser *L. luscinia* (pale shade) in the breeding season (northern) and in winter (southern). Black areas show where the ranges overlap (redrawn from Cramp, 1988).

are limited to particular altitudes. Other centers of endemism are less obviously isolated and it is often postulated that they are places that formed habitat islands in the past as climatic changes fragmented their currently more continuous habitat. Since it is wooded rather than open habitats that tend to have been broken into such habitat islands, this fits with 71% of the restricted-range species being forest birds (and another 13% being birds of dry woodland and scrub). Alternatively, rivers and hills may be sufficient barriers to dispersal that they create the boundaries to centers of endemism.

D. Biogeographic Regions

Information on birds was important in the original identification of the six biogeographic regions of the earth—regions characterized by individual peculiarities of their faunas and floras as a result of long periods of separation in the past. Some bird families are found in several regions but others are confined to just one. The level of endemism in a region corresponds to the extent of its past isolation.

The least isolated regions, in that they form two great landmasses that have been joined by a broad land bridge (at the Bering Straits) for much of their recent history, are the Nearctic (North America) and Palearctic (Europe, most of Asia, and North Africa). The first has no endemic bird families and the second only one; many families occur in both, probably as a result of invasions from the Palearctic to the Nearctic via the Bering land bridge. The climate is markedly seasonal in large parts of each of these regions and they support comparatively few species (approximately 750 in the Nearctic and approximately 1100 in the Palearctic), many of which are migrant insectivores that move south in winter.

TABLE VIII

Percentage of Bird Species That Are Ecologically Separated from Congeneric Species Primarily by Feeding, Habitat, or Geographic Range (i.e., Having Largely Separate but Adjoining Ranges) or by Having No Contact[a]

	Parus species (%)		Trans-Saharan passerine migrants (%)		European birds (%)		
	N. America	Europe	In Europe	In Africa	Passerines	Other land birds	Waders and waterbirds
Feeding	4	43	10	2	19	43	48
Habitat	11	32	62	23	55	20	18
Range	29	8	8	64	7	11	9
No contact	55	17	18		18	14	19
Unknown	—	—	2	10	1	11	7

[a] Data from Lack (1971).

Many species were probably lost from these regions as a result of deterioration of the climate during the Pleistocene. Subsequently, the Caribbean Sea, the Mediterranean Sea, and the Sahara desert, although crossed by many migrants, have been partial barriers to recolonization from Africa and South America.

The Oriental region (the Indian subcontinent and Southeast Asia) has only 1 endemic bird family, although it remains somewhat isolated from the Palearctic by mountains. It has approximately 1000 species. Much of its avifauna originated largely in the Palearctic, although it has some affinities with the Ethiopian region (sub-Saharan Africa). The latter has 13 endemic families among its approximately 1500 species. It receives many Palearctic migrants during the northern winter and some of its resident species may have ancestrally originated in the Palearctic. The richness of forest birds in Africa is less than that of Central and South America, perhaps reflecting the disruption to African forests caused by dry periods in the Pliocene and Pleistocene.

The Australian region has a distinct avifauna—16 endemic families and approximately 1000 endemic species among the approximately 1600 bird species living there. Many of these belong to the Corvida, a major division of the Passeriformes revealed by DNA analysis, which appears to have originated in Australia. New Guinea and other nearby islands belong to this region, the somewhat unclear dividing line between the Australian and the Oriental regions lying between Lombok and Bali and between Celebes and Borneo, along a deep strait that has probably always separated these regions even when sea levels have been much lower than they are today. The ancestors of some Australian species have arrived from the north across this barrier, but others are derived from the south, at a time when Australia lay with the other Southern Hemisphere continents in the great landmass of Gondwanaland.

The most distinctive and most species-rich region is the Nearctic (South America), inhabited by 31 endemic families and more than 3000 species living in a variety of habitats ranging from the tropical to the far south temperate and from sea level to extensive mountains. It has a comparatively small proportion of oscines among its avifauna and seems to have had a long evolutionary independence from the rest of the world, although its passerines may be largely derived from North America.

E. Evolutionary Convergence

When unrelated forms occupy similar habitats in different parts of the world, they may evolve similar characteristics in adaptation to those habitats. Such evolutionary convergence between unrelated forms is widespread among birds: for example, the great auk *A. impennis* and penguins (large, flightless seabirds using wings to swim under water); the little auk *Plautus alle* and diving petrels *Pelecanoides* (small seabirds that plunge into the depths to catch planktonic prey, using wings to both swim and fly); the New World vultures (related to storks) and the Old World vultures (related to hawks); the hornbills and toucans of the Old and New Worlds, respectively (both with long bills for reaching fruit in the canopy, with secondary use of the decorated bill for social signaling); and the sunbirds and the hummingbirds, also of the Old and New Worlds, respectively (small, brightly colored nectar feeders, with specialized tongues). More generally, the Australian Corvida include many groups of birds that have been considered in the past to be more closely related to various groups of passerines elsewhere in the world than to each other. Such conclusions depended on similarities that they shared with these other groups as a result of occupying similar ecological niches and thus being molded by similar selection pressures. These similarities are now recognized as convergent.

The extent of morphological convergence among whole communities has rarely been examined. A comparison of birds from Mediterranean communities in France, California, and Chile showed that they were no more similar to each other than each was to bird communities from a temperate part of France. Perhaps the environmental differences between temperate and Mediterranean France are too slight to promote significant morphological divergence.

V. DIVERSITY ON ISLANDS

A. The Colonization of Islands by Birds

Seabirds abound on islands partly because they are safer places to nest than mainlands (generally having fewer predators) but also because the oceans present no barriers to seabirds. In contrast, land and freshwater birds are less diverse on islands than on mainlands, although they have been much more successful in reaching islands than have other nonmarine vertebrates. Continental islands, which were once joined to larger landmasses before being separated by continental drift or sea-level rise, may retain some of their original species (perhaps much modified) but the evidence indicates that the more ancient of them have been colonized by many other species across the sea. Recent genetic evidence that kiwis are related to emus suggests, for

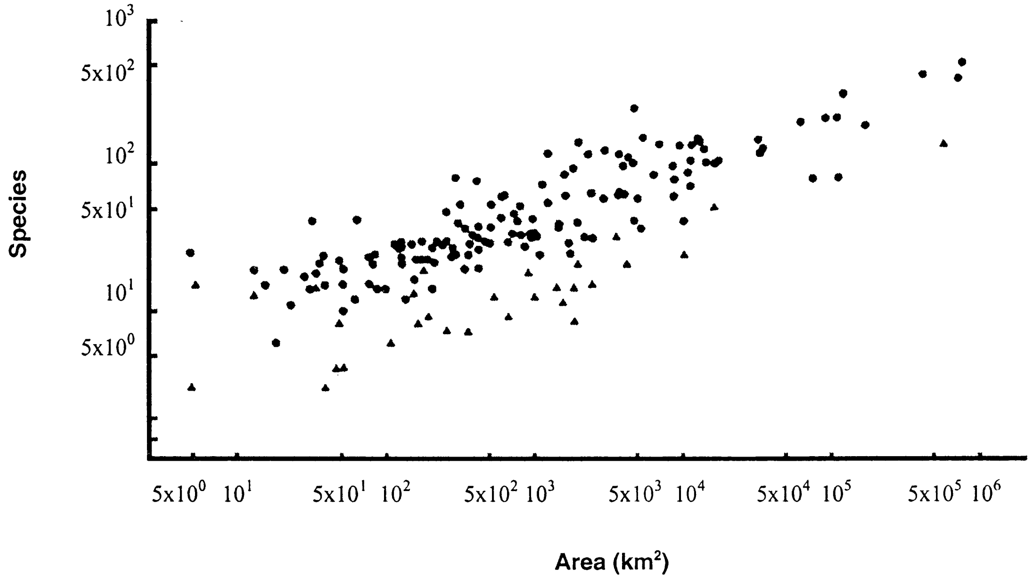


FIGURE 18 Relationship between the number of species on individual islands in the warmer seas and their areas. Circles represent islands more than 300 km from the next largest landmass or in the Hawaiian or Galapagos archipelagos (redrawn from Williamson, 1981).

example, that their ancestors may have arrived from the north, perhaps by flying over the Tasman Sea. (Moa, in contrast, probably originate from an original Gondwanaland stock.) Oceanic islands, never joined to a mainland, must have been colonized over the sea. Vagrant individuals of a variety of species are commonly observed even on the most remote islands. The efficacy of such colonization is shown by the fact that those islands of Krakataua that were devastated by volcanic eruption in 1883 held as full a complement of species as undisturbed islands within less than 100 years. As another example, if one allows for differences in area (small islands tend to hold fewer species), oceanic islands are as rich in bird species as continental islands in the Gulf of California.

Even for birds, however, isolation appears to result in substantial loss of diversity (Fig. 18). Islands 250 km from New Guinea have only about half the number of bird species expected from their size, and those more than 800 km away have fewer than 10%. However, is this because birds do not reach remote islands? Perhaps they get there but do not find rich enough habitats (because a sufficient wealth of plants and other animals has not arrived). The answer undoubtedly depends on the species and the island in question. Studies that focus on individual species, rather than just on counts of species number, show clearly that species differ markedly in colonizing ability. J. M. Diamond conducted extensive analyses of the bird faunas of the many satellite islands of New Guinea and found, in addition to simple distance effects, that islands that were connected by land bridges during periods of low sea level in the Pleistocene had more bird species than those that had not had such connections. Some species, apparently poor at crossing water, are found only on land-bridge islands, even though the never-connected islands have the same habitats available. Diamond divided the birds according to apparent colonizing ability. At one extreme are "supertramps," often seen as vagrants, found on the most isolated islands that few other species have colonized, living at high densities, breeding rapidly, and having unspecialized habitat requirements. At the other extreme are species that are found only on islands that many other species have also reached; they are habitat specialists and live at low densities. When they reach an island these specialists appear to reduce the resources to such a degree that the generalist supertramps cannot survive; the latter are therefore absent from species-rich islands. Thus, the supertramp dove *Ptilinopus solomonensis,* found on many small islands of the Bismarck Archipelago, is replaced on New Britain and New Ireland by three congeners—one in mountains, one in lowland forests, and the third in open lowland habitats.

B. Evolution of Island Birds

Island populations are genetically somewhat isolated (sometimes wholly isolated) from the mainland populations from which they sprang. The environment on the island is likely to differ from that on the mainland both physically and biotically (having fewer competitors and predators). As a result, there may be rapid evolutionary adjustment. Flightlessness has already been mentioned. Niches may be broadened. Thus, in most of western Europe, treecreepers *Certhia familiaris* are largely confined to conifers (and generally to high altitudes) by the presence of the closely related *C. brachydactyla,* which is a broad-leaved specialist, but in Britain and Ireland (and also in far eastern Europe and Asia), from which *C. brachydactyla* is absent, they inhabit woods of all sorts. Where studies of the relationship between bird species diversity and foliage height diversity have been conducted on islands, it has been found that the birds appear to divide the foliage into fewer layers than on the mainland; individual species have broader niches when there are fewer competitors. Another evolutionary change in some groups, such as ducks, is that island forms may lack distinctive male breeding plumages, either because there is less chance of hybridization when related species are absent or because demographic changes lead to behavioral modifications.

Island forms are often recognized as subspecies as a result of such changes; they may even become distinct species.

The chaffinch *Fringilla coelebs* is widespread in Europe and North Africa, inhabiting both deciduous and coniferous woods. It has the same habits on the Canary Islands of Hierra and La Palma, but on Tenerife and Gran Canaria it is restricted to deciduous woods; coniferous woods on those islands are occupied by the related *F. teydea.* These islands appear to have been colonized by *F. coelebs,* which then, adapting to the widespread coniferous woodlands there, evolved so much (into *F. teydea*) that when another *F. coelebs* invasion occurred the two species not only failed to interbreed but also each had a habitat to which it was better suited than the other and in which it could therefore persist. Such species pairs are known elsewhere. They represent a rare case of island bird communities being more diverse than those of mainlands.

Studying the birds of the Lesser Antilles, R. E. Ricklefs and G. W. Cox modified the idea of a "taxon cycle," earlier propounded by E. O. Wilson in respect of Australasian ants to describe the distributional and evolutionary changes when a species colonizes an archipelago. *Molothrus bonairiensis,* progressively colonizing the southern islands (from South America), represents an early stage of such a cycle; *Tyrannus dominicensis,* found throughout the islands, represents the next; and then *Loxigilla noctis,* with distinct subspecies on some islands. Finally, some of the islands have actually lost the colonizing species, either through chance extinction or through competition from later invaders: *Dendroica adelaidae,* restricted to St. Lucia and Barbuda, is one such species.

Isolated archipelagos, with very depauperate faunas but diverse habitats and with some isolation between islands, are hotbeds of evolution for those species that do arrive. The adaptive radiation of the Hawaiian honeycreepers (Drepanididae) is a well-known example. Comprising 23 extant species (more than half the endemic bird species on Hawaii) and at least 8 extinct ones, they are derived from a single fringillid ancestor but with remarkably divergent feeding adaptations: typical finch beaks for seed-eating, parrot-like beaks, thin insect-eating beaks, and beaks elongated up to one-third the body length for sucking nectar. The akiapolaau *Hemignathus wilsoni* has a stout lower mandible with which it chips into decaying wood, holding out of the way its long, curved upper mandible, which it then uses to probe for insects in the exposed galleries. Darwin's finches (Geospizinae) are even better known. The diversity of the Galapagos species of this subfamily is in sharp contrast to the existence of just 1 species on Cocos Island, about 700 km northeast. The latter, although it has a diversity of habitats, has not presented the opportunities for speciation that the many separate islands of the Galapagos have done; as a result, the Cocos finch is behaviorally flexible, utilizing a wide range of foraging opportunities.

C. The Species–Area Relationship in Birds

It is well-known that larger islands tend to hold more species (see Fig. 18). Birds have been used in many studies of the species–area relationship because they have reached most islands and their distributions are well-known. An especially steep relationship has been observed for islands in a Minnesota lake, apparently because the smaller islands studied were just too small (<0.5 ha) to support even single territories of many species. However, this is unlikely to account for the relationship observed in many studies of much larger islands. This has been widely interpreted in terms of the MacArthur–Wilson theory of island biogeography: As more species become established on an island, the rate at which new species colonize is bound to decrease (fewer of the arrivals are of new species); as the number of established species increases, the number going extinct also increases (there are more to go extinct); and at some point the colonizations and extinctions therefore reach an equilibrium. Since extinctions are less likely on larger islands (because populations are larger), the equilibrium number of species will be greater on larger islands. On well-studied islands, such as those around the British coast, there are certainly frequent colonizations and extinctions. These British islands are close to the source of colonists and are small, so this result is scarcely surprising, but it illustrates the dynamic nature of island avifaunas. The data also show that species differ in their probabilities of extinction; therefore, species-specific approaches are needed for a complete understanding.

The slope of the species–area relationship tends to be flatter for more isolated island groups because these are colonized only by the species with better powers of dispersal, which can quickly reoccupy islands from which they have gone extinct. On Bahamian islands, the slope is shallower for migrant species than for residents, the former being better colonists (and the slope is steeper for lizards, which are poor colonizers).

For species late in the taxon cycle, where extinctions have begun to take effect, the slope is steeper than for species early in the cycle, as the equilibrium theory would predict. Conversely, the slope was found to flatten in a group of Finnish islands following reduction in disturbance by man; this had previously led to many extinctions, especially on small islands.

Although these observations are in agreement with the equilibrium theory, they do not rule out the importance of habitat. David Lack argued that larger islands held more bird species because they were more diverse, both physically and in terms of floristic richness. Statistical analyses designed to untangle the effects of habitat diversity and those of area *per se* found the former to be the chief cause of the species–area relationship in birds on the Isles of Scilly and on montane "islands" in the American Great Basin. Species richness of birds on the Lesser Antilles correlates independently with both habitat and area; in divergent contrast, bats correlate only with area and butterflies, reptiles, and amphibians only with habitat. It has been suggested that the difference is the result of birds (and bats *a fortiori*) having lower population densities than the other animals and therefore being more vulnerable to chance extinction. Similarly, the low species richness of the most isolated Bahamian islands seems to depend on both their isolation (many birds do not get there) and their habitat poverty (many arrivals find nowhere to live). Studies of forest fragments, which show the same sort of species–area relationship, indicate that part of the reason for the small number of species in small fragments is that they lack those species that are only found deep in forests.

Competition and resources mold island avifaunas. There is evidence that the species–area relationship is less steep for individual trophic groups than for whole avifaunas, which is in agreement with the idea that competition within the group reduces the ability of new species to colonize when an island already has many species or increases the likelihood of a new colonization leading to the extinction of an established competitor. Competition may also explain the paradoxical finding that migrant birds are more species rich on more remote Bahamian islands; the migrants, it is hypothesized, can reach all the islands but are less likely to establish on islands with residents already established (i.e., the less isolated islands).

D. Fragmentation as a Cause of Extinction

Nature has conducted experiments that reveal the importance of the area of islands in determining extinction rates and thus the numbers of species present at equilibrium. Thus, islands that were connected to New Guinea as a result of lower sea levels approximately 10,000 years ago now have fewer species than expected if they were part of mainland New Guinea (although they still have more species than do islands of comparable size that were never connected, suggesting that equilibrium has not yet been reached). Three New Guinean islands that were part of a single island approximately 10,000 years ago have fewer species than that single island would be expected to have, given its estimated size (although again they have more than expected for islands of their current size). The number of species that each has apparently lost is inversely proportional to its area, as the equilibrium theory predicts. The same is true for a suite of islands off the north coast of South America and the coast of Central America.

Man formed Barro Colorado Island during the construction of the Panama Canal early in the twentieth century. It has subsequently lost 50–60 species. Some of the losses may be the result of increased predation because man no longer hunts small and medium predators. It is striking, however, that the species that have gone extinct are those whose numbers appear to be particularly affected by the occasional periods of very dry weather that occur in the region. Average low population densities, in contrast, seem not to be a good predictor of extinction from Barro Colorado: As long as a population does not fluctuate too much, it can persist even if its average numbers are small.

Forests in most parts of the world have been fragmented by man and numerous studies have shown that fragmentation has resulted in the loss of species. This is not just the effect of random extinction, however, because it particularly involves the loss of forest-interior specialists. Indeed, direct observation in the highly fragmented woodlands of eastern England confirms that many more woodland species pass through small woodlots than stay to breed, showing that they would have no difficulty in colonizing such places if they were suitable. Blue tits *P. caeruleus* and great tits *P. major* breeding in such small woods produce fewer young, partly as a result of breeding later in the season, than do the same species breeding in larger woods. This may be because the small woods provide less shelter from the weather or because they lack sufficiently rich vegetation to produce adequate insect food for the birds. Whatever the cause, these fragmented woods support fewer birds than would a solid block of forest of the same total area.

In North America, Neotropical migrants also tend to avoid fragmented forests, much more than do resident species or temperate migrants. The fragmentation of forests may be the reason why so many Neotropical migrants have declined in numbers in recent decades.

E. Introduced Species as Colonists: A Challenge to the Equilibrium Theory?

Many birds have been introduced outside their native ranges and have flourished in their new homes. For example, 34 introduced land and freshwater species are established on New Zealand. Many of these have colonized offshore islands—up to 10 on a single island. This does not mean that those islands were not in equilibrium before these species arrived. What has happened is that the pool of species available to colonize the islands has been increased, raising the colonization rate and thus setting a new equilibrium.

If New Zealand were capable of absorbing these species, does this mean that it was not at equilibrium? With only 60 resident, nonmarine indigenous species it was certainly depauperate, so one could argue that ecological niches were vacant. However, in undertaking the introductions, man was enhancing the colonization rate, thus shifting the equilibrium upward. Furthermore, many native species have been driven close to extinction in New Zealand, with some surviving only because of special protection measures. Although habitat loss and the devastations of introduced predators have been major causes of such near losses, competition from the introduced species may also have played a part.

VI. HUMAN CAUSES OF THE LOSS OF AVIAN BIODIVERSITY

A. Natural Causes of the Loss of Avian Biodiversity

The paleontological record shows that many birds have become extinct as part of the natural flux of life on Earth. The recent extinctions of an unnamed megapode on the Kermadec Islands (1876) and of the San Benedicto rock wren *Salpinctes obsoletus exsul* (1952) were also natural, resulting from volcanic eruptions. That element of biodiversity that is made up of the differences between regions in the species that inhabit them must also have been eroded by natural colonizations. These natural processes, however, are slow—slow enough for extinctions to be approximately balanced by the origins of new species. Man has accelerated them manyfold, causing serious losses of avian biodiversity.

B. The History of Bird Extinctions

It is difficult to be sure whether a species has gone extinct, especially if it is scarce and lives in remote places. The Fiji petrel *Pterodroma macgillivrayi,* for example, was known from but a single specimen taken in 1855 until a second was found in 1984. With that caveat, Table IX shows the avian extinctions recorded in the past 400 years. These represent more than 1% of known birds, a proportion similar to that for mammals but otherwise greater than that for any other group of animals or plants.

Two things stand out from Table IX. First, extinctions peaked between 1850 and 1950 but have still continued at a high rate. Second, approximately 90% of them have been extinctions of island species, even though only 20% of bird species live on islands. One-third of the birds present on Hawaii 200 years ago are now extinct and two-thirds of the remainder are endangered. Two-thirds of the historically indigenous birds of the Mascarenes have been lost and more than 40% of the rest are endangered. Island species are partly vulnerable because their ranges are so small and some of the extinctions on continents have been of species

TABLE IX

Chronology of Known Extinctions of Bird Species Since 1600[a]

	1601–1650	1651–1700	1701–1750	1751–1800	1801–1850	1851–1900	1901–1950	1951–present	Total
Continental	0	0	0	0	2	2	5	2	11
Island	4	8	5	8	14	31	30	15	115
Total	4	8	5	8	16	33	35	17	126

[a] From Newton (1998).

with particularly small ranges. Because birds that occur in few places tend to be uncommon even where they do occur (Figs. 5 and 6), they are especially vulnerable.

Hunting has caused many extinctions. As recently as 1945, Japanese troops eliminated the Wake Island rail *Rallus wakensis.* The case of the passenger pigeon *Ectopistes migratorius* is classic. Once numbered in hundreds of millions, it occurred in huge flocks that "darkened the skies." Although habitat loss may have contributed to its decline, there is little doubt that it was extinguished by massive commercial hunting. Such uncontrolled exploitation contrasts with situations in which a community has been able to control the hunting, such as was the case on St. Kilda, where seabirds were a major resource whose exploitation was regulated. The great auk *A. impennis* and Dodo *Raphus cucullatus,* extinguished by visitors to their uninhabited island homes, provide a clear contrast and also illustrate other aspects of the vulnerability of many island birds: They were large, flightless, and had limited antipredator responses, thus making them easy and profitable to hunt. They probably had low reproductive rates and therefore were incapable of sustaining high levels of hunting.

The introduction of alien species has been a major cause of island extinctions. Worldwide, there is a clear correlation between the number of indigenous birds that have become extinct in a locality and the number of avian introductions established there. Some introduced species, especially mammals such as rats, cats, dogs, mongooses, and pigs, are important as predators. In the early 1960s, black rats *Rattus rattus* eliminated the last known population of the bush wren *Xenicus longipes* within 3 years of reaching Big South Cape Island, New Zealand, along with the island's populations of four other birds. In 1894, the lighthouse keeper's cat caught 15 Stephen Island wrens *Xeniscus lyalli;* these were the first specimens discovered—and also the last. Introduced monkeys are important predators at the nests of endemic birds on Mauritius. Only one introduced snake is known to have caused serious problems, but its effects have been dramatic: The brown tree snake *Boiga irregularis* seems to have arrived on Guam in the 1940s as an accidental hitchhiker. By 1986 it had extinguished six of the indigenous forest birds and reduced the other four to populations of less than 100; of these, the endemic Guam rail *Rallus owstoni* had been reduced to 100 in 1983 and it was extinct in the wild by 1987, although it survives in captivity. In contrast, none of the seven introduced birds on Guam has been obviously affected by the snake: They have effective antipredator behavior and live in habitats in which the snake is scarce. (Only one of the eight native species in such habitats has gone extinct.) It is noteworthy that few of the species extinguished by introduced predators lived on islands with indigenous rats or land crabs, which would have caused them to maintain their antipredator responses.

Introduced birds may out-compete natives, especially in the modified habitats that man has created on most islands since they may be adapted to such habitats in their own native ranges. Thus, the familiar birds in suburban and agricultural habitats of New Zealand are the familiar birds in suburban and agricultural habitats of western Europe. Competition may be genetic as well as ecological: Part of the cause of the extinction of the Atitlán grebe *Podilymbus gigas* may have been hybridization with the introduced pied-billed grebe *P. podiceps.*

The loss of natural habitats has been another major cause of extinctions. Much of it has been deliberate (the clearance of land for agriculture or human settlement), but much has been an accidental by-product of the introduction of alien herbivores, especially to islands where the native plants have evolved in the absence of such animals. Thus, the rabbit *Oryctolagus cuniculus* destroyed much of the vegetation on Laysan in the 20 years following its introduction in 1903; three of the five species of land birds disappeared with the plants. On some islands, soil erosion, following deforestation, has removed nest sites for burrowing seabirds.

Finally, the impact of introduced diseases is largely unknown but native Hawaiian species, previously free of avian malaria, have been infected via introduced birds.

C. The Recent Prehistory of Bird Extinctions

It is well established that there were mass extinctions of large mammals following the colonization of continents by man the hunter. Vulnerable birds may have been extinguished at the same time, as may have birds dependent on the mammals. Fossils show the Californian condor *Gymnogyps californianus* once to have been widespread across North America, its restriction to the west presumably consequent on the reduction in numbers of large mammals brought about by man, dating from approximately 11,000 years ago. It now survives only through a captive breeding program, its final historical collapse also being a by-product of human activity—in this case, poisoning by the lead in carcasses abandoned by hunters and by poisons used to kill coyotes *Canis latrans*.

Colonization of islands mostly occurred later. In the Pacific, people had gone as far as the Solomon Islands by 30,000 years ago but did not reach western Melanesia and Micronesia until 4000 years ago, moving then through Fiji and Samoa, the Marquesas, Hawaii, and Easter Island, with New Zealand only colonized about 1000 A.D. European colonists followed in the sixteenth–eighteenth centuries. Some of these human colonizations eventually failed, perhaps because of their dramatic impact on the islands' ecosystems. The early colonists hunted the animals they found, generally introduced alien species, and usually cleared forests. Their activities certainly led to loss of bird populations: On Hawaii, 60 endemic species are known only from bones; at just one site in New Zealand, 29 endemics are known only from bones; on Easter Island, bones of about 30 species of seabirds have been found from the early years of human occupation but only 1 is now found there (with about 10 others surviving on offshore islets) and all the indigenous land birds seem to have died out following the total deforestation of the island. D. W. Steadman reached the following conclusions about Polynesian land birds:

> (i) The ranges of most living species are much smaller today than they were at the time of first human contact. (ii) Few volant species are naturally endemic to only one or two islands. (iii) Most species have become extinct in the past 3000 years. (iv) Most or all islands supported one to four endemic species of flightless rails, virtually all now extinct. (v) In many cases, individual islands supported two or three species within a genus, unlike the situation today. (vi) At least four formerly widespread genera now are gone from East Polynesia. (vii) Although modern distributions of Polynesian land birds continue to be analyzed as if they were natural, they do not furnish unbiased data for proposing or testing ecological models. (viii) Although some of the range losses of extant species could be restored with conservation efforts, we are centuries too late to preserve any true likeness to the original Polynesian avifauna.

On the basis of the fossil evidence, Steadman suggests that the 800 Polynesian islands inhabited by man have lost an average of 10 species or populations of birds, including 2 or 3 species of flightless rails endemic to each. Other authors have suggested that, because of the incompleteness of the fossil record, even these data may be underestimates. Whatever the details, the avifauna that survived to be further devastated by Europeans, Americans, and Japanese was a small part of that which the earlier colonists found. The marked deficiency of seabirds in the tropics, although probably partly a result of the low productivity of tropical oceans, may be a legacy of this destruction.

D. Current Threats to Birds

On the basis that they have undergone (or appear likely soon to undergo) rapid declines, that they have small and declining populations, that they leave small and declining ranges, or that they have very small populations, 11% of the species of birds in the world are classed as "threatened." The percentage for mammals is similar; no other group of animals reaches this level of threat. Forest birds make up 65% of this total, and those of scrub and of wetlands each comprise 9%.

Most of the endangered species are tropical. At least for the New World, this is not just because there are more species in the tropics because the proportion of the avifauna that is threatened is also highest in the tropics (Fig. 19). This appears to be in part because the tropics hold many species with restricted ranges. (Throughout the world, about one-fourth of the threatened species are endangered simply by having very

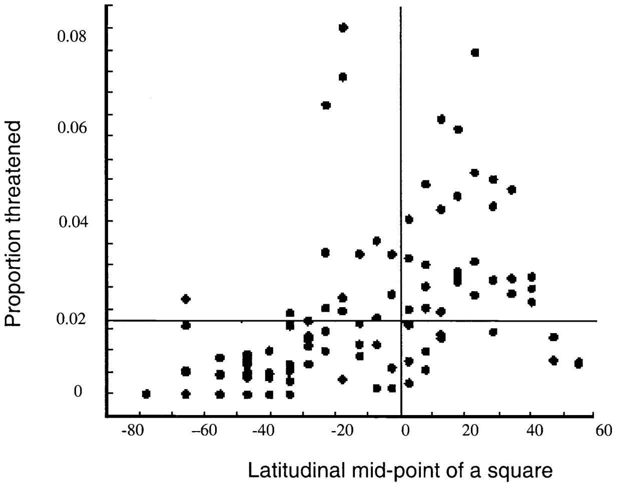


FIGURE 19 Proportion of New World birds classed as threatened, in relation to latitude (negative values are in the Northern Hemisphere). The horizontal line is the mean proportion threatened (redrawn from Gaston and Blackburn, 1996).

small populations or ranges.) In the New World, it is also because of the intensity of tropical deforestation. Of course, areas that hold relatively few threatened species may do so because those species that were especially vulnerable to the impact of man have already gone extinct; this may explain why there are fewer threatened species north of the equator in the Americas than in the south (Fig. 19), given the longer history of human interference in the north.

Habitat loss, especially the loss of tropical broadleaved forest, is a threat to about half of the world's endangered birds. It is caused not only by the expansion of agriculture but also by mining and the construction of dams. In many parts of the world, afforestation with exotic species is a major problem: Being exotics, the introduced species offer homes to only some of the native animals. Furthermore, they are managed in ways that reduce the floristic and structural diversity of the plantations. Drainage of wetlands is a major problem in many regions, although one that can often be relatively easily reversed if there is the will to do so. About 8% of the world's endangered species are threatened by hunting and capture for the pet trade. In Europe, more than 30% of the species that have declined in recent years are thought to be affected by hunting or by related persecution (such as the killing of raptors that are viewed to be competitors with man). Much hunting is now recreational rather than for subsistence, but it is often more destructive than formerly because there are more people, because more of them have the leisure time to engage in hunting, because there is better technology (mist nets, guns, and motor transport), and because more people can afford the technology. Coastal habitats in sunny places are steadily being destroyed by the development of tourist facilities. Even where beaches are left in a natural state, birds breeding on them suffer from increased disturbance.

In Europe, the greatest threat, affecting more than 40% of declining species, is the intensification of agriculture. Birds of the steppes are threatened by irrigation, whereas those of wet meadows are threatened by drainage. Fewer resources are available for wildlife when hay meadows are converted to silage and when stocking rates on pastures are increased. Insecticides and herbicides remove much potential bird food from crops. Plant breeding, the use of pre-emergence herbicides, and more effective machinery may allow autumn cultivation in areas where fields used to be left in stubble over winter, such that granivores could feed on grain spilled at harvest and on the seeds of weeds. Harvesting is more efficient. Fundamentally, more of the primary productivity is being harvested, so less is available for wildlife. Most of these changes depend on technological developments but their adoption has been promoted by systems of agricultural support, such as the Common Agricultural Policy of the European Union (EU), that concentrate on production rather than also taking account of social and conservation needs. Expansion of the EU to include countries of Eastern Europe, where agriculture is still relatively less intensive, is a major threat to birds and other wildlife.

Further agricultural changes may flow not only from technology and policy but also from global climate change. This will also affect natural habitats. Although vegetation and wildlife have adapted to the natural climate changes of the past 10,000 years, often apparently within decades, there are concerns that the changes now occurring will be substantially more disruptive because of their speed.

E. Reducing the Threats

Species have been pulled back from the brink. Use of DDT reduced the Mauritius kestrel to 4 birds in 1974 but 20 years later, through a program of captive breeding and reintroduction, there were more than 200. There have been several similar cases, but they have required very intensive care. More extensive measures are needed if all threatened birds are to be conserved.

There are two broad approaches to such conservation. One is simply to control all of the threats listed previously, which would have a widespread impact. Some actions would be neither politically very difficult nor economically very costly, such as banning deliberate introductions, controlling accidental introductions, and eliminating introduced species that have become established. Most, however, would entail a direct conflict between the benefits to wildlife and the material wealth of burgeoning human populations.

The other remedy is to secure individual sites as biodiversity reserves. This would also be expensive because, although 20% of bird species occupy just 2% of the land surface of Earth, it would take a much larger area to include the other 80%. (Furthermore, because hot spots of diversity for different taxa do not closely coincide, even greater areas would be needed if more than birds were to be considered.) In any case, diversity hot spots are peculiar places; conserving them might maintain many species but it would not conserve the natural diversity of ecosystems. There is also the practical difficulty of ensuring that the correct locations are chosen. For example, after decades of trying to build up numbers of the red kite *Milvus milvus* in the Welsh fastnesses to which the British population had become restricted by nineteenth century persecution, conservationists concluded that it was probably in Wales not because the most suitable habitat was there but just because it had not been shot out there; populations subsequently introduced to eastern England and Scotland have flourished, indicating that the Welsh birds were probably living in relatively poor conditions.

F. Other Reductions in Local Diversity

Extinction is a particularly obvious way in which biodiversity is diminished but is merely the tip of an iceberg. For every species extinguished by man, there are many others whose numbers and ranges are substantially reduced. (There are also a few whose numbers have been much increased by human activity, such as the house sparrow *Paaser domesticus* and starling *S. vulgaris.*) Changes in numbers are less easy to assess than extinctions and little is known about them in many parts of the world. There are good data, however, for much of North America, where it is clear that recent decades have seen major declines of many Neotropical migrants. It is not clear to what extent this is the result of habitat loss in the wintering areas or of fragmentation of woodlands on the breeding grounds. In Europe, changes in agriculture have been closely followed by declines in the populations of farmland bird specialists (Fig. 20). Studies of individual species are beginning to reveal the causes of these losses, which are different in detail for different species but all fundamentally related to agricultural intensification.

It is specious to argue that such changes are but a reflection of the natural flux of populations. Although systematic counts are not available for most species, except for recent years, it is clear from the extensive historical literature that British farmland birds have probably never before suffered losses on the scale of those recorded in the past few decades, even during the periods of enclosure and other intensive habitat

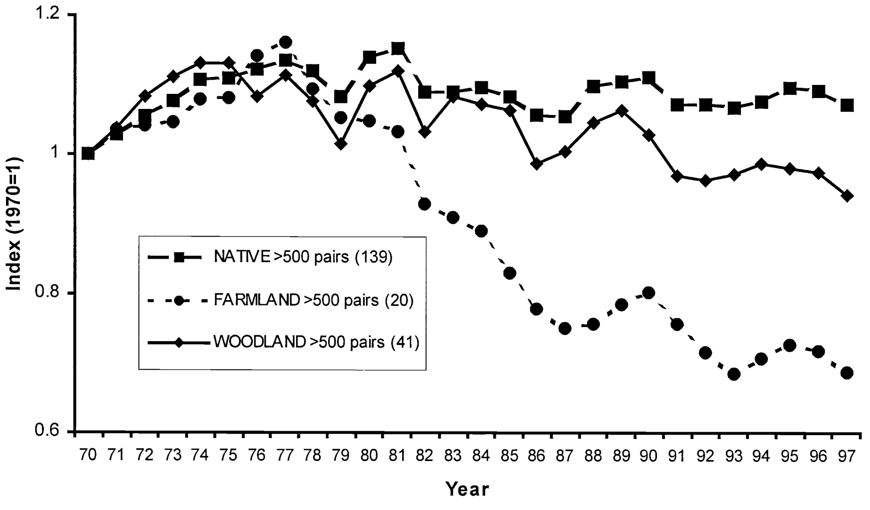


FIGURE 20 Index of bird populations in the United Kingdom based on geometric means of indices for individual species. Introduced species and those with fewer than 500 pairs are excluded (reproduced with permission of the British Trust for Ornithology and the Royal Society for the Protection of Birds).

modification near the end of the eighteenth century and the development of high agriculture in the nineteenth century. The argument that the species characteristic of farmland are much more common than they were before much of western Europe was cleared of its forests is correct but incomplete. Forest clearance led to open-country birds replacing woodland birds, but the farmland birds now being lost are not being replaced. As a result, the total biomass of nonmarine birds breeding in Great Britain declined between 1968 and 1988 by 10% (39% if the pheasant *Phasianus colchicus,* which is widely and increasingly stocked for shooting, is omitted).

G. Losses of Geographical Diversity

Part of the diversity of life on Earth is the difference between places in their animal and plant communities. Introductions not only endanger biodiversity at the local scale (see Sections VI,B and VI,C) but also, however harmless they are to the indigenous fauna and flora, lessen biodiversity by reducing the difference between places.

More than 130 species of birds have been naturalized outside their native ranges, i.e., they have been introduced and have established breeding populations. As a result, the range of some species has increased dramatically. The pheasant *Phasianus colchicus,* native in only a band stretching from the Caucasus to Japan, now occurs in most of Europe, temperate North America, and New Zealand, with scattered populations in Australia and various islands.

Sport is a major reason for introductions (even in regions with plenty of native game birds) and one-fifth of established introductions are of pheasants (Phasianidae). Some species have been introduced for food (such as the feral pigeon *Columba livia*), many to remind settlers of "home" (such as European songbirds across the globe), and some because they are beautiful (e.g., peafowl *Pavo cristatus*). Some have been introduced as agents of biological control, such as the house sparrows *P. domesticus* transported to the United States in the nineteenth century to control defoliating moths. The Chimango caracara *Milvago chimango* was introduced to Easter Island as a scavenger. Conservationists have occasionally introduced threatened species beyond their native range because the latter was too badly affected by habitat destruction or by introduced predators or sometimes, perhaps less justifiably, because the native range was considered unsafely small.

Many species have established themselves following accidental escape from captivity; therefore, wildfowl, parrots, doves, and small seed eaters make up a high proportion of naturalized species. In the summer of 1991, Britain (which, like most of western Europe, has only one native species of goose) held established breeding populations of four other geese; 10 other species were at large, 2 of them recorded breeding, and there were 18 different sorts of hybrid geese identified. Some of these avicultural species may be deliberately released because their owners tire of caring for them or wish to see them free-flying. Similarly, farmed species may be released if markets collapse: This is the origin of the Australian population of the ostrich *Struthio camellus.*

Species that are closely associated with man, such as house sparrows and house crows *Corvus splendens,* have reached some places by riding on ships.

Introductions of birds have been less damaging than those of mammals and fishes. Nonetheless, each one of them represents a direct reduction in the diversity of life throughout the world.

See Also the Following Articles

AMPHIBIANS, BIODIVERSITY OF • ENDANGERED BIRDS • FISH, BIODIVERSITY OF • MAMMALS, BIODIVERSITY OF • MIGRATION • REPTILES, BIODIVERSITY OF

Bibliography

Blackburn, T. M., and Gaston, K. J. (1994). The distribution of body sizes of the world's bird species. *Oikos* **70**, 127–130.

Blackburn, T. M., and Gaston, K. J. (1996a). Spatial patterns in the body sizes of bird species in the New World. *Oikos* **77**, 436–446.

Blackburn, T. M., and Gaston, K. J. (1996b). Spatial patterns in the species richness of birds in the New World. *Ecography* **19**, 369–376.

Blondel, J. (1991). Birds in biological isolates. In *Bird Population Studies* (C. M. Perrins, J.-D. Lebreton, and G. J. M. Hirons, Eds.), pp. 45–72. Oxford Univ. Press, Oxford.

Campbell, B., and Lack, E. (Eds.) (1985). *A Dictionary of Birds*. Poyser, Calton, UK.

Collar, N. J., Crosby, M. J., and Stattersfied, A. J. (1994). *Birds to Watch 2. The World List of Threatened Birds*. BirdLife Int., Cambridge, UK.

Cramp, S. (Ed.) (1988). *The Birds of the Western Palearctic*, Vol. **5**. Oxford Univ. Press, Oxford.

Fuller, R. J. (1982). *Bird Habitats in Britain*. Poyser, Calton, UK.

Gaston, K. J., and Blackburn, T. M. (1996). The spatial distribution of threatened species: macroscales and New World birds. *Proc. R. Soc. London B* **263**, 235–240.

Gaston, K. J., and Blackburn, T. M. (2000). *Pattern and Process in Macroecology*. Blackwell Science, Oxford.

Grant, P. (1999). *Ecology and Evolution of Darwin's Finches*, 2nd ed. Princeton Univ. Press, Princeton, NJ.

Gregory, R. D., Greenwood, J. J. D., and Hagemeijer, E. J. M. (1999). The EBCC atlas of European breeding birds: A contribution to science and conservation. *Biol. Conserv. Fauna* **102**, 38–49.

Greenwood, J. J. D., Gregory, R. D., Harris, S., Morris, P. A., and Yalden, D. W. (1996). Relations between abundance, body size and species number in British birds and mammals. *Philos. Trans. R. Soc. London B* **351**, 265–278.

Hochberg, M. E., Clobert, J., and Barbault, R. (Eds.) (1996). *Aspects of the Genesis and Maintenance of Biological Diversity*. Oxford Univ. Press, Oxford.

Jones, P. H. (1972). Succession in breeding bird populations of sample Welsh oak woods. *Br. Birds* **65**, 291–299.

Lack, D. (1971). *Ecological Isolation in Birds*. Blackwell Scientific Publications, Oxford.

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S., and Watt, A. D. (1998). Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature* **391**, 72–76.

Morse, D. H. (1975). Ecological aspects of adaptive radiation in birds. *Biol. Rev. Cambridge Philos. Soc.* **50**, 167–214.

Moss, D. (1978). Diversity of woodland songbirds populations. *J. Anim. Ecol.* **47**, 521–527.

Newton, I. (1998). *Population Limitation in Birds*. Academic Press, San Diego.

Newton, I., and Dale, L. C. (1997). Effects of seasonal migration on the latitudinal distribution of west Palearctic bird species. *J. Biogeogr.* **24**, 781–789.

Nilsson, S. G. (1979). Density and species richness of some forest bird communities in south Sweden. *Oikos* **33**, 392–401.

Ralph, C. J. (1985). Habitat association patterns of forests and steppe birds of northern Patagonia, Argentina. *Condor* **87**, 471–483.

Schall, J. J., and Pianka, E. R. (1978). Geographical trends in numbers of species. *Science* **201**, 679–686.

Schluter, D., and Ricklefs, R. E. (1993). *Convergence and the regional component of bird species diversity*. In *Species Diversity in Ecological Communities* (R. E. Ricklefs, and D. Schluter, Eds.), pp. 230–240. Univ. of Chicago Press, Chicago.

Sibley, G. C., and Ahlquist, J. E. (1990). *Phylogeny and Classification of Birds*. Yale Univ. Press, New Haven, CT.

Terborgh, J. (1980). Vertical stratification of a neotropical forest bird community. *Acta XVII Congr. Int. Ornithol.* **2**, 1005–1012.

Wiens, J. A. (1989). *The Ecology of Bird Communities: Volume 1. Foundations and Patterns*. Cambridge Univ. Press, Cambridge, UK.

Williamson, M. (1981). *Island Populations*. Oxford Univ. Press, Oxford.

BOREAL FOREST ECOSYSTEMS

Roy Turkington
University of British Columbia

GLOSSARY

boreal A term derived from the Greek word for north reflecting the fact that this vegetation type, the boreal forest, occurs only in the Northern Hemisphere.

permafrost That part of the soil profile that is permanently frozen and which forms a barrier to water drainage often resulting in a wet surface condition.

podzol A soil type highly characteristic of boreal forests and developed as a consequence of podzolization.

podzolization The process of acid leaching whereby clay, organic particles, and mineral ions (primarily iron and aluminum) are carried downwards and deposited in the B soil horizon, leaving an impoverished and leached A horizon. This occurs as a consequence of low temperatures and precipitation in excess of the needs of evapotranspiration.

taiga Mostly used as a synonym for boreal forest, but more precisely it is a Russian word applied to Eurasian conifer forests described as damp and almost impenetrable. It is also defined as a coniferous forest with no admixture of nonconiferous tree species except *Betula* and *Populus*.

THE BOREAL FOREST, or Taiga, is a broad northern circumpolar belt of predominantly coniferous forests. It has many lakes and vast areas are covered by forested wetlands and bogs; bogs and meadows of varying sizes are found throughout the boreal zone. It is the world's largest vegetation type, occupying approximately 14 million km^2, or 8% of the world's continental ecosystems. It accounts for approximately 12% of the world's biomass. This article will describe the distribution of the boreal forest and the major factors influencing this distribution, such as climate, soils, and fire. The apparent uniformity of the forests throughout their range masks a real diversity that is evident at regional and local scales. This diversity, particularly of plants but also of animals, will be described.

I. GEOGRAPHY

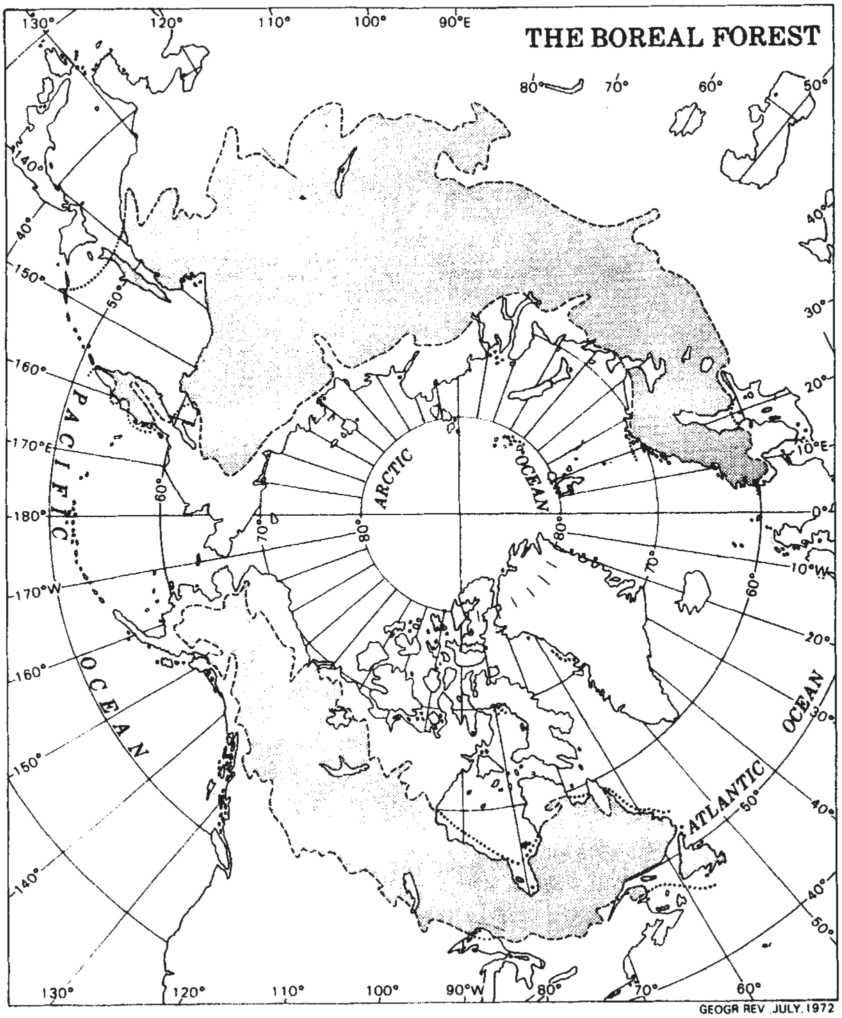
The boreal forests are confined to the Northern Hemisphere. In North America the forest is a continuous vegetation belt stretching across the continent and spanning more than 10° latitude. It's northern limit is defined by transition to the treeless tundra; this limit

Encyclopedia of Biodiversity, Volume 1

extends from about 68°N in Alaska to 58°N on the west coast of Hudson Bay, and it reaches the Labrador coast at about 58°N (Larsen, 1980). In Fennoscandia it ranges from 56°N to 69°N (Esseen *et al.*, 1997), with the northern limit fringing the northern coast of Norway. Across Russia, except for a small border, the northern boundary more or less follows the Arctic coast. In parts of Siberia the boundary is up to 500 km inland (Larsen, 1980) (Fig. 1). The southern limits of the boreal forest are more difficult to define because the boundaries are rarely sharp. In more oceanic areas the forest is bounded to the south by broad-leaved deciduous forest, and in continental areas transition is to parkland, dry grasslands, and semideserts. For example, in western Canada the southern transition is to subalpine forest, in central Canada it is to prairie grasslands, and in easterly regions it is to mixed deciduous forest.

II. CLIMATE

There are strong relationships between the climate and the soils of the boreal regions. The characteristic nature of the climate to a large extent dictates the nature of the soils, including the permafrost, and these ultimately determine the plants and animals that live there.



GEOGR REV JULY 1972

FIGURE 1 The circumpolar range of the boreal forest. About two-thirds of the area is in Eurasia. The sector in eastern Canada lies farthest from the North Pole (from Hare and Ritchie, 1972. Reprinted by kind permission of the author and the American Geographical Society).

A. General Limits

The climate of the boreal forest has been extensively documented by, for example, Hare and Ritchie (1972), Larsen (1980), Elliott-Fisk (1988), and Bonan and Shugart (1989). The boreal forest grows where winters are too long and summers too short to support temperate forests. This typically occurs when the growing season is less than 6 months and the frost-free period less than 4 months (or fewer than 4 months with temperatures higher than 10°C). The boundary between boreal forest and tundra corresponds approximately with (i) the line south of which the temperature is higher than 6°C for 4.5 months (Hare and Ritchie, 1972), (ii) the position of the July 13°C isotherm with marked departures in regions with montane or oceanic influences (Larsen, 1980), or (iii) where there are less than 30 days with a daily mean temperature higher than 10°C and where the cold season lasts 8 months (Walter, 1973).

B. Temperature

In more general terms, Elliott-Fisk (1988) describes the climate as cool, humid microthermal, with very cold winters of 7–9 months allowing persistence of snow cover during all but the brief (3 or 4 months), relatively cool summer season. For more than 6 months of the year, the mean temperature is below 0°C and net radiation is negative. Maximum summer temperatures generally reach the low 20s and winter minimums in the −50s. However, yearly variation can be extreme such as in Verkhoyansk (30°C in summer and −70°C in winter), but these extremes are moderated near the coast, e.g., in Umeå, Sweden (Fig. 2).

C. Precipitation

Generally, the boreal zone is characterized by having a high proportion of the annual precipitation falling as snow. Throughout most of the boreal forest annual precipitation is low, varying between 250 and 1000 mm (Fig. 2). There are extremes, however, with Fort Yukon in Alaska and Verkhoyansk in Siberia each recording 180 mm annually. Some areas of western Norway with up to 2000 mm of precipitation annually have been classified as boreal rain forest. Because of low temperatures and short growing seasons, evaporation rates are low and drought is uncommon. When drought does occur, forest fires can ravage vast areas of boreal forest.

III. SOILS

Boreal forest soils are typically low in fertility and acidic, with a thin A horizon. The most characteristic soils are podzols. However, podzols occur in a wide range of climates, not only in boreal regions, and not all boreal regions are underlain by podzols. Climate, vegetation type, chemical composition of the substrate material, and topography are the major environmental influences that produce the typical boreal podzol. The combination of low temperatures and low pH impede decomposition processes and slow the rate of soil development. The soil surface may be covered by a mat of spruce needles up to 3–7 kg/m^2. This mat of acid, partly decomposed plant material is the mor litter layer. The slow but gradual decomposition of this layer continually releases a supply of organic acids that contribute to the leaching of organic particles and mineral ions (primarily iron and aluminum) from the surface soils

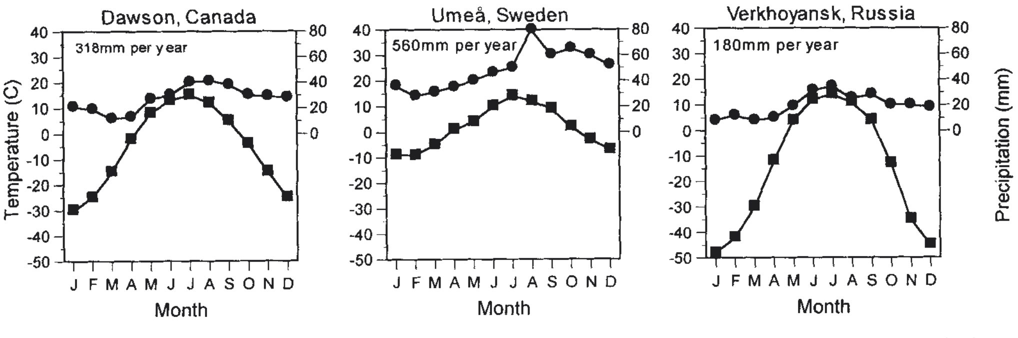


FIGURE 2 The annual variation of average temperature (■) and precipitation (●) of three locations in the boreal forest zone; Dawson in the Yukon Territories, Canada (altitude, 324 m), Umeå, Sweden (7 m), and Verkhoyansk, Russia (137 m).

and to the weathering of the parent material. In addition, soluble materials such as sodium, potassium, and calcium are washed out of the soil by water movement. As a consequence, the surface soils are relatively infertile and high in silica. In many cases in which the process of podzolization is prominent, the upper soil layers are gray or whitish in color. Beneath the leached layer is a zone in which materials leached downward by water accumulate, chiefly iron–humus complexes. These deposits may be cemented into a hardpan, sometimes thick and strong enough to prevent root penetration to the lower soils. In the more extreme boreal forest climates the subsoil is permanently frozen (the permafrost). The combination of nutrients being largely tied up in the litter layer, an infertile A horizon, hardpans and permafrost results in most boreal forest trees having a shallow root network.

IV. FIRE

Individual trees or local stands may be killed by windstorms, landslides, snow avalanches, erosion, ice storms, flooding, or insect attack, but the most important natural disturbance in the boreal forest is fire. Also, although the boreal forest is relatively impoverished floristically, much of the floristic diversity and large-scale vegetation mosaics are directly attributable to repeated burning. Longer periods of unseasonably drier and warmer weather lead to a greater probability of fire, and it is the few major fires during extreme fire years that account for the majority of forests burned; this inevitably has enormous consequences for the wildlife that inhabit the forests, particularly those unable to take refuge in water bodies. In northwestern Canada and Alaska, between 60 and 80% of all fires are less than 5 ha in area, but in extreme fire years individual fires can burn up to 200,000 ha. Because fires often burn all or most of the forest floor, they influence organic matter accumulation, soil temperature, and soil moisture, and through these they impact major ecosystem processes such as nutrient cycling, energy flow, and productivity. The natural cycle of fire frequency in boreal forests of North America ranges from an average of 50–200 years to up to 500 years in wetter parts of eastern Canada. Fire frequencies in northern Sweden average from 110 to 155 years but may be as high as 270 years. Fire frequency varies considerably, both within and between regions, resulting in a mosaic landscape with different forest patches in various stages of succession. Factors contributing to the probability of a fire include tree species composition, stand structure, soil conditions, the amount and moisture content of the fuel, exposure, topography, and climate; there is much disagreement regarding whether the probability of a fire increases with stand age. Pine forests burn, on average, twice as frequently as spruce forests. Local fire chronologies indicate that fire frequency is highest in the interior, more continental areas due to a higher frequency of thunderstorms and lightening strikes; most natural fires are caused by lightening strikes. In the boreal forest of Fennoscandia there is apparently a north to south gradient with a generally lower fire frequency toward the north.

V. VEGETATION

The boreal forest is the most continuous and extensive forest in the world. The North American and Eurasian forests are remarkably uniform in their appearance throughout their range, both in their physiognomic structure and in species composition. Typically, there is a simple canopy layer (15–20 m high) in which numerical dominance is maintained nearly everywhere by coniferous tree species belonging to four genera: spruce (*Picea*), pine (*Pinus*), fir (*Abies*), and larch (*Larix*) (Fig. 3). Species of juniper (*Juniperus*), cedar (*Thuja*), and hemlock (*Tsuga*) also occur. Next is a shrub layer (typically 1 or 2 m tall) supporting mainly broad-leaved deciduous species that are also frequently present as successional components of the forest; deciduous species rarely achieve dominance, except in some postfire successions and in the mountain birch forest in Fennoscandia. Generally, the deciduous species belong to the genera willow (*Salix*), birch (*Betula*), poplar (*Populus*), and alder (*Alnus*). The herb layer typically is poorly developed but it is enmeshed in a well-developed ground layer of mosses, liverworts, and lichens. Despite the uniformity of appearance, the tree species are unique to a particular continent, i.e., there are no circumboreal trees (Table I). The species making up the shrub layer and ground layer may be more wide ranging, and many of the mosses and lichens are circumboreal. Bryophytes and lichens are typically a more common component of the forest floor vegetation than vascular plants. Bryophytes usually dominate on mesic and moist sites and their diversity in boreal forest is higher than in most temperate or tropical forests. Lichens also contribute significantly to plant diversity in boreal forests, especially on drier and northern sites, and particularly in the ground layer of pine heaths and on rocky ground.

Because there are few species of trees, the boreal forest gives an impression of monotony, but this is

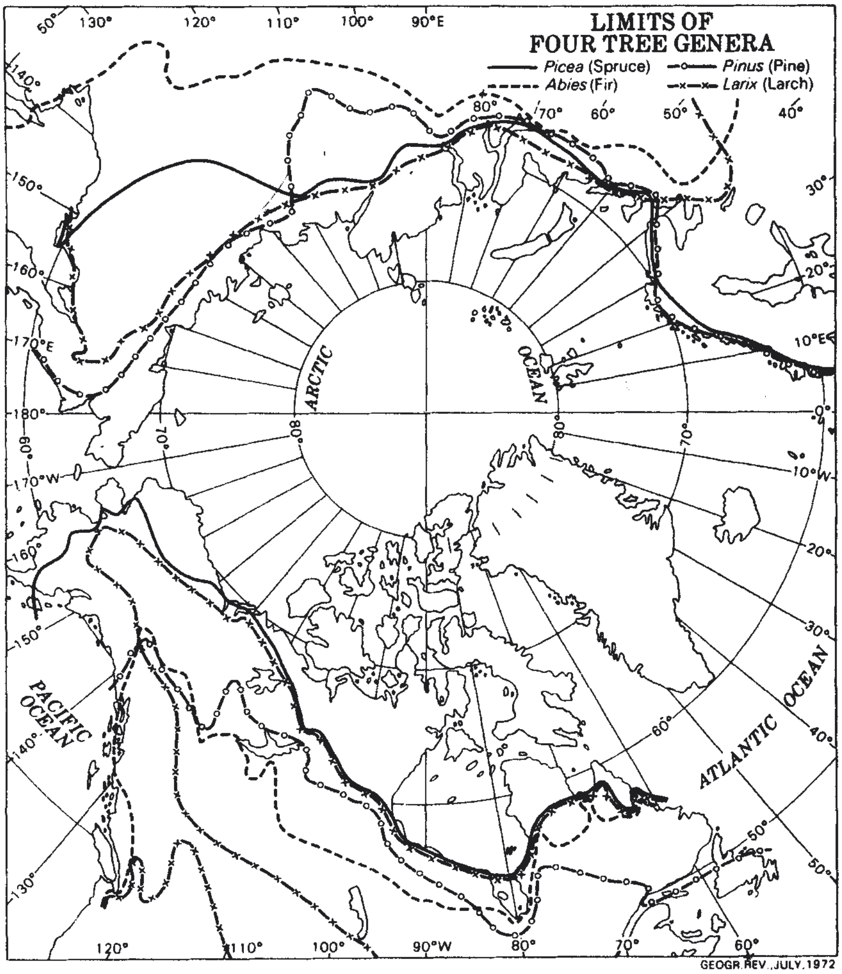


FIGURE 3 Limits of *Picea, Abies, Pinus,* and *Larix,* without regard to species. Limits are greatly generalized, being envelope curves around numerous disjunct areas. In particular, the central Alaskan distribution of *Larix* is probably disjunct from the Mackenzie Valley distribution (from Hare and Ritchie, 1972. Reprinted by kind permission of the author and the American Geographical Society).

misleading even though they are floristically impoverished compared to most other vegetation formations of the earth. Nevertheless, these forests are a complex mosaic of different plant communities, and they caused Cumming *et al.* (1996) to title their paper "Boreal Mixedwood Forests May Have No 'Representative' Areas." Some of the tree species are widely distributed (e.g., *Picea glauca* and *P. mariana*), but many other boreal species have more limited ranges, resulting in local and regional changes to forest composition. Throughout the North American boreal forest, a vegetation group dominated by *P. glauca* (sometimes codominated by *Abies balsamea, A. bifolia, Betula papyrifera, or Populus tremuloides*) is typically the climax community in relatively dry areas. A second group dominated by *P. mariana* occurs where soils are wetter, such as in bogs and muskegs. *Picea mariana* is not strictly characteristic of wet areas because it also flourishes after fire. In the European forests the Scots pine (*Pinus sylvestris*) is frequently dominant on drier soils and the Norway spruce (*Picea abies*) occurs on moister sites.

Although there are no abrupt transitions in the vegetation, changes in the dominant tree species, in the subdominant shrubs, and in the herbaceous layers can be detected. Various attempts have been made to survey and describe latitudinal and longitudinal gradients, based primarily on climatic distinctions, and to describe the associated vegetation assemblages. For example,

TABLE I

Major Species of Boreal Forest Trees by Longitudinal Section[a]

Genus	North America, (55°W–160°W)	Northern Europe (5°E–40°E)	Western Siberia (40°E–120°E)	Eastern Siberia (120°E–170°E)
Conifers				
Picea (spruce)	*glauca* *mariana*	*excelsa**	*obovata*	*obovata* *jezoënsis*
Abies (fir)	*balsamea*		*sibirica*	*nephrolepis*
Pinus (pine)	*banksiana*	*sylvestris*	*sibirica* *sylvestris*	*sylvestris* *pumila* *cembra*
Larix (larch)	*laricina*		*sibirica* *sukachzewski*	*dehurica*
Hardwoods				
Populus (popular)	*tremuloides* *balsamifera*	*tremula*	*tremula*	*tremula* *suaveolens*
Betula (birch)	*papyrifera* *kenaica*	*pubescens* *verrucosa** *kusmisscheffi*	*verrucosa** *pubescens*	*ermani*
Alnus (alder)	*tenuifolia* *crispa* *rugosa*	*incana*	*fruticosa*	*fruticosa*
Salix (willow)	*Salix* species	*Salix* species	*Salix* species	*Salix* species

[a] Certain species of nonboreal type range into southern parts of the forest. In North America, for example, these include the white and red pines (*Pinus strobus* and *P. resinosa*), *Thuja occidentalis*, and species from *Ulmus*, *Fraxinus*, and *Acer*. In Europe and Asia, species of *Quercus*, *Tilia*, and *Fraxinus* also have a limited boreal range. The taxonomies of *Betula Alnus* and *Salix* are highly confused, and this table is hence only an approximate guide.

*[Author's note: *Picea excelsa* should be *Picea abies* (Norway or European spruce), and *Betula verrucosa* should be *Betula pendula* (European birch)]. From Hare and Ritchie, 1972. Reprinted by kind permission of the author and the American Geographical Society.

Larsen (1980) described 7 regions in the boreal forest of North America, and Rowe (1972) described 35 regions. However, rather than describe each of these, I focus on some of the recent classifications.

A. A Latitudinal Classification

Hare and Ritchie (1972) recognized three latitudinal zonal divisions of the boreal forest in North America: a southern closed forest, a central lichen woodland, and a northern forest–tundra ecotone.

1. Closed Forests

The closed forest dominates vast areas of the southern boreal forest, and it occurs on a wide range of soil types and topographies. These forests have a closed crown with a moist, deeply shaded floor. The spruce–feathermoss community is characteristic of this zone, with either white or black spruce dominant. The *P. mariana*–feathermoss forests have a fairly uniform, moderately dense tree stratum, with an almost continuous ground cover of bryophytes. In contrast, the *P. glauca* forests are more irregular and open, having an understory of broad-leaved shrubs and herbs and a patchy bryophyte distribution. The most common moss in the black spruce forests is *Pleurozium schreberi*, whereas *Hylocomium splendens* is more characteristic of the white spruce and mixed woodlands.

2. Lichen Woodland

These forests are more open with a discontinuous shrub layer and abundant lichens. The transition from the southern closed forests to these lichen woodlands is usually smooth or mosaic-like but can sometimes be quite abrupt as in northern Manitoba and Saskatchewan. In some cases, a Western *Stereocaulon paschale*- and an Eastern *Cladonia stellaris*-dominated woodland have been identified. *Picea mariana* and *P. glauca* are the dominant trees in these forests, with *P. glauca* declining to the north. Because these forests are more open, light penetration to ground level is much greater than in closed forests. Although these trees may be as tall (but less dense) as those in closed forests, the

TABLE II

Major Dominant, Codominant, or Abundant Species Identified in the Three Major Longitudinal Sections in Peinado *et al.*'s (1988) and La Roi's (1967) Classifications of North American Boreal Forests

Classification	
Peinado *et al.* (1998)	La Roi (1967)
The western section (northeastern British Columbia, southern Yukon, and Alaska)	
Abies lasiocarpa	*Populus/Salix/Shepherdia*
Pinus contorta var. *latifolia*	
Picea engelmanii	
The central section (eastern British Columbia, Alberta, Saskatchewan, and Manitoba)	
Abies lasiocarpa	*Lonicera/Rubus pubescens/Lathyrus ochroleucus*
Abies bifolia	*Lonicera/Vaccinium vitis-idaea-Geocaulon*
Abies balsamea	
The eastern section (all forests east of Lake Winnipeg, Manitoba)	
Abies balsamea	*Kalmia-Picea mariana*
Pinus banksiana	*Abies balsamea*

additional light promotes branching to ground level. The additional light often promotes a discontinuous layer of heath plants such as crowberry (*Empetrum nigrum*) and bilberry (*Vaccinium myrtillus*).

3. Forest–Tundra Ecotone

Here, there are scattered and isolated trees, often deformed or prostrate, in a tundra landscape. Although this type of forest is located north of the limit of continuous forest, it is still considered boreal forest. The ecotone is more than 300 km wide in Quebec, up to 225 km wide in Central Canada, and it narrows at both its eastern and western ends.

B. A Longitudinal Classification

There is a remarkable similarity between the vegetation of the boreal forests in eastern North America and that in eastern Asia, with many identical or closely allied genera and sometimes species. These forests contain many different species, whereas those of the Euro-Siberian region contain few. Of all the tree species in the North American boreal forest, only *P. glauca* extends from the Bering Straits across Alaska and Canada to Newfoundland. *Picea mariana,* usually found only on poor or wet soils, is found at the timberline toward the Arctic, and *Larix laricina* is found in the continental regions. Only two species, *P. sylvestris* and *P. abies,* are of any real importance in the boreal zone of Europe. Only in eastern regions of Europe is *P. abies* replaced by the closely related *Picea obovata,* whereas additional species are being added to the forest (*Abies sibirica, Larix sibirica,* and *Pinus sibirica;* Table I). Moving east, spruce gradually declines from the forest until it is entirely absent in eastern Siberia.

Peinado *et al.* (1998) analyzed the vegetation of the North American boreal forests. They identified three major sections (Table II) and classified them as eight major groups (Table III). The three sections coincide well with La Roi's (1967) classification for the same regions.

Qian *et al.* (1998) examined longitudinal patterns of plant diversity in the North American boreal forests, focusing specifically on the southern closed forests. The central section has a higher species and genera diversity than the western and eastern sections (Table IV). White spruce forests are always more diverse than black spruce forests; this is a reflection of the diversity of herbaceous plants and bryophytes and not the diversity of woody plants (Fig. 4). The diversity of white spruce forests is rather similar between western and eastern sections, but the diversity of black spruce forests is much higher in the west than in the east (Table IV). The diversity of bryophytes is remarkably consistent across the continent, but again there are more bryophytes in white spruce than in black spruce forests (Fig. 4).

VI. ANIMALS

The boreal forest is home to many animals. It is the winter home of the migratory caribou and reindeer and the permanent home of many others. The wolf and lynx are the major predators of the boreal forest. Some of

TABLE III

Major Dominant, Codominant, or Abundant Species Identified in the Eight Groups in Peinado *et al.*'s (1998) Classification of North American Boreal Forests

Group	Western	Central	Eastern
I	*Picea glauca* and/or *Pinus contorta* var. *latifolia* *Populus tremuloides*	*Abies bifolia* *Abies lasiocarpa*	*Picea glauca* *Picea mariana* *Abies balsamea* *Betula papyrifera*
II	*Picea mariana*	*Abies balsamea*	*Picea rubens* *Abies balsamea* *Betula alleghaniensis*
III			*Alnus rugosa*
IV			*Pinus banksiana*

the best examples of population cycles in animals are described from the boreal forest regions; for example, lynx (*Lynx canadensis*), snowshoe hares (*Lepus americanus*), arctic ground squirrels (*Spermophylus parryi*), red squirrel (*Tamiasciurus hudsonius*), and boreal red-backed vole (*Clethrionomys rutilus*) in northern Canada and microtine rodents, owls, capercaillie (*Tetrao urogallus*), black grouse (*Tetrao tetrix*), mountain hare (*Lepus timidus*), and red fox (*Vulpes vulpes*) in Eurasia. The causal relationships of these cycles have not been fully explained but many of the North American examples are synchronized with the snowshoe hare cycle, and Eurasian examples are synchronized with microtine rodent cycles.

In North America, other inhabitants of the boreal forest include moose, black bear, grizzly bear, deer, wolverine, coyote, marten, beaver, porcupine, sable, voles of the genus *Microtus*, chipmunks, shrews, and

TABLE IV

Mean Number of Species and Genera in White Spruce and Black Spruce-Dominated Ecosystems in Three Geographic Sectors of the Boreal Forest of North America[a]

	Western North America		Central North America		Eastern North America	
	Species	Genera	Species	Genera	Species	Genera
White spruce ecosystems						
No. of plots	13		10		11	
Woody plants	13	11	18	14	13	10
Herbaceous plants	28	25	33	30	25	24
Bryophytes	23	19	25	20	26	22
Vascular plants	41	35	51	44	38	33
All plants	65	54	77	63	65	55
Black spruce ecosystems						
No. of plots	7		10		9	
Woody plants	14	10	14	11	12	11
Herbaceous plants	23	19	28	24	19	16
Bryophytes	15	13	15	13	14	12
Vascular plants	37	30	41	36	31	27
All plants	53	42	57	49	46	39

[a] The boreal forest can be latitudinally divided into three subzones: a wide zone of closed forest in the south, a narrow band of forest–tundra ecotone in the north, and a lichen woodland in the middle. These data cover almost the entire longitudinal range but focus on the southern closed forest (modified and adapted from Qian *et al.*, 1998, by kind permission of the author and Kluwer Academic Publishers).

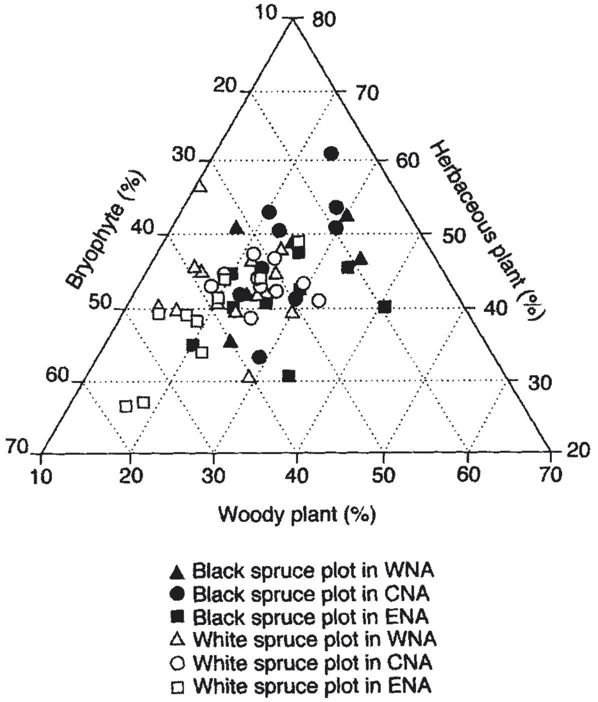


FIGURE 4 Triangular ordination showing sample plots in the white spruce and black spruce ecosystems of the North American boreal forest. Each point represents a single sample plot of 9 ha (300 × 300 m). Thirty-four of the plots were sampled in *Picea glauca* forest and 26 in *Picea mariana*. CNA, central North America; ENA, eastern North America; WNA, western North America (from Qian *et al.*, 1998. Reprinted by kind permission of the author and Kluwer Academic Publishers).

bats. Typical animals of the boreal forest vary slightly more from location to location than do the plants. The moose which browse on willow, birch, alder, and water plants, and the beaver which feeds on aspen, are widespread. Many birds also inhabit the boreal forest; for example, great horned owl, goshawk, spruce grouse, ruffed grouse, nuthatchers, juncos, and warblers. Brown bears inhabit the boreal forest in Eurasia.

In the boreal zone of Eurasia, the diversity of mammalian herbivores is highest in the interior of the continent and declines to the east. Across Eurasia, species richness of mammalian herbivores is positively correlated to warm climate, the number of hardwood species, and the area of the boreal forest. Across North America, species richness of mammalian herbivores increases as the length of the growing season and the number of coniferous tree species increase (Fig. 5). Given this information, it appears that indirect measures of primary productivity as well as the number of tree species can accurately predict species richness of mammalian herbivores. Bird diversity decreases from west to east across both the North American boreal forests and the Eurasian boreal forests. In Fennoscandia the diversity of forest birds decreases northwards; in Finland, this occurs only in pine forests and not in spruce.

It seems that the boreal forests of Canada, and possibly Russia, differ from those in northern Fennoscandia in that small herbivore biomasses reach much higher levels and are dominated by species of hare rather than voles. In addition, the densities of many fewer species in the boreal forests of Canada are correlated with the dominant herbivore relative to the situation in Fennoscandia. In Fennoscandia, where voles fluctuate, their changes in density are strongly correlated with that of hares, and those of two predominant alternative prey, red squirrels and voles, were not.

Tree death and decaying wood provide a variety of habitats for an enormous number of invertebrates. For example, in Sweden approximately 1000 species of beetle are dependent on dead trees. The most diverse fauna on snags is found during the first 2 years after the tree has died. Spruce logs have a more diverse invertebrate fauna than pine, but many invertebrates can inhabit both. Four typical stages in the succession of invertebrates on spruce logs in boreal forests have been described. Initial colonization is by bark beetles and other primary cambial eaters along with their associated parasitoids, predators, and detritivores. Subsequent stages have been described in detail by Esseen *et al.* (1997).

VII. SUCCESSION

Many studies of the boreal forest have shown a remarkably close fit between the plant community and the environment, especially with reference to moisture conditions and fertility level. This relationship is not static and communities are always in a process of change. The sequence of primary successional change is typically described as beginning in fens and then progressing as follows:

> Fen → swamp → bog → muskeg → invading trees → forest

Fens develop under alkaline conditions on a peat substrate usually with standing water. Bogs are specialized communities of shrubs and herbs growing on a wet, acidic, peat substrate. The sequence becomes drier and more mineral poor as muskeg develops, and this is followed by tree invasion.

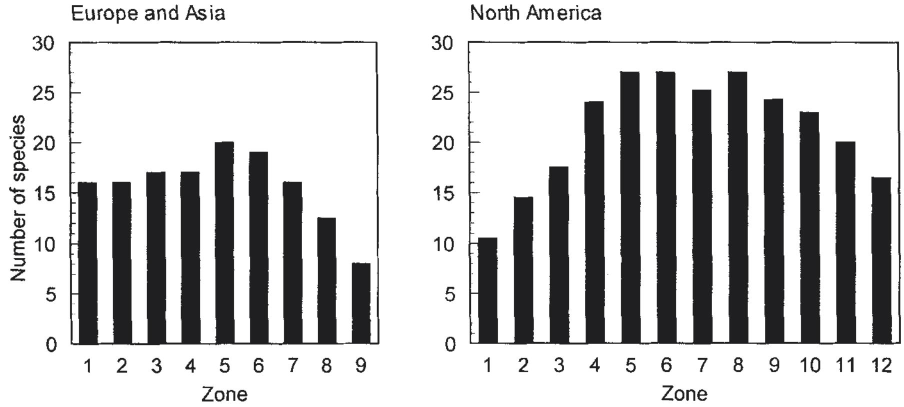


FIGURE 5 The number of mammalian herbivore species along a gradient from west to east (zones 1–9 across Europe and Asia and zones 1–12 across North America). Note the decline in species number as the continents meet at the Bering Strait (adapted from Danell *et al.*, 1996).

The usual sequence of secondary succession is described as

Disturbance → annual plants → perennial herbs and shrubs → larger shrubs and trees → forests

However, these descriptive sequences mask a dynamic system under the influence of frequent disturbances, and the boreal forest is better considered as a disturbance forest. This means that the concept of the vegetation climax is probably not applicable because most of the tree species in these forests are incapable of self-perpetuation in climax state. They are instead, adapted to fire, and the boreal forest is usually maintained in a pre-climax state by frequent fires. Fire is the primary disturbance factor interrupting the successional process, and the frequency and intensity of fires will determine the nature and stage of succession. Local topographic, soil, moisture, and microclimate conditions will determine the local fire regime, and if the local regional environment is patchy so also will be the frequency of fires. The consequence will be a vegetation mosaic, with each patch being at a different stage of succession and representing a different degree of recovery since the last fire. In this sense, the boreal forest climax is not a stand dominated by white or black spruce, but is a mosaic in which all stages of a successional or regeneration cycle are represented, i.e., the climax is a mosaic of successional patches. It has been argued that if fires could be excluded from a lichen woodland for at least 200 years, then it is likely, that a closed-forest spruce–feathermoss would develop; however, fires usually occur more frequently and so the lichen woodland is perpetuated.

Nevertheless, immediately following a fire some species are characteristic pioneer species. Among herbaceous plants, *Epilobium angustifolium* and *E. latifolia* (fireweeds) are the most obvious. Density and cover of *Salix, Betula,* and *Alnus* are typically high in early pioneer stages and greatly decrease in later stages when they are intolerant of the shady conditions imposed by the establishing *Picea, Abies, Pinus,* and *Populus.* Many questions remain concerning the nature of succession in the boreal forest. There are many regional differences, and local site conditions exert considerable influence of the successional process—both on its path and on its "final" form.

VIII. CONSERVATION

In both Canada and Russia, vast areas of boreal forest are being cut annually, and the situation is worse in Fennoscandia. Although wood cutting is the biggest danger to biodiversity, there are additional threats from mining, pollution, road building and dam construction. Clear-cutting vast areas of forest has profoundly altered the landscape structure in northern Sweden and in central Canada, resulting in habitat loss, habitat alteration, and fragmentation. Clearly, this will have mostly negative influences on both animal and plant biodiversity, even though some species will benefit in the short-term, i.e., weedy species such as *Epilobium angustifolium, Deschampsia flexuosa,* and *Calamagrostis purpurea* that are adapted to fire and other disturbances. Few data are available on the long-term effects of clear-cutting in the boreal forest, but studies from other forests suggest that herb communities do not recover in logging cycles of 40–150 years. Rotational cutting and clear-cutting will inevitably influence, or eliminate, natural fire regimes and will lead to changes in tree species composition and forest structure, a general reduction in stage age, and reduced input from coarse woody debris (Esseen *et al.*, 1997), all of which will have profound impacts on the natural biodiversity of the boreal forest.

Biodiversity must be preserved at all scales from the genetic variation within a population to heterogeneity occurring at the landscape level. There are two major approaches to preserving biodiversity in these areas: (i) sustainable management and (ii) reserves of natural areas. The first approach is more applicable to areas that are already subject to intense management such as Fennoscandia, whereas the second approach can be applied where vast, relatively untouched areas still exist such as in parts of Canada and Russia.

A. Sustainable Management

To maintain biodiversity, we must also preserve or simulate the processes, mostly natural disturbances, that produced the heterogeneity in the first place. Microscale heterogeneity may be enhanced by gap disturbance and by coarse woody debris. Larger scale heterogeneity may be enhanced by fire or insect outbreak. Angelstam (1998) proposed a conceptual model as a guide to the maintenance and restoration of ecologically sustainable boreal forests. The model is based on the hypothesis that self-sustained forest ecosystems can be created, or recreated, and their biodiversity developed if forest management can simulate the composition and structure of boreal forest landscapes. This is done by introducing and maintaining disturbances that lead to naturally dynamic spatial and temporal patterns of forest regeneration. The model has already been implemented in planning systems of many large Swedish forest enterprises.

B. Reserves of Natural Areas

Reserves alone are not sufficient to conserve biodiversity in forests, but any such initiative for maintaining biodiversity in the boreal forest should include at least three components:

1. Large areas representative of southern closed-canopy forest, lichen woodland, the forest–tundra ecotone and each of the longitudinal elements within these must be protected.
2. In selecting these areas, attention must be given to the animal inhabitants, their abundance or rarity, and their migratory behavior, if any.
3. The decision-making process should consider the "floating reserve" strategy (Cumming *et al.*, 1996) in which portions of a protected area could be periodically replaced in response to aging of components, unexpected large-scale disturbance, or refinements in conservation objectives.

Although much of the world's boreal forest still remains intact with little impact from man, the impact from man has been immense in many areas. In the new millennium, there will continue to be an increasing demand on the world's resources, including the boreal forests. Indeed, man's impact has already been substantial in Fennoscandia (less than 5% of the boreal forest remains in a natural or seminatural state) and is increasing in the North American forests. It is tempting to argue that these forests are the world's largest vegetation type, occupying vast areas (approximately 14 million km^2) and account for approximately 12% of the world's biomass, so the incursion of man is likely to have little impact. As described in this article, although the boreal forest is less diverse than most of the other world's vegetation types, it is nevertheless a complex mosaic of patches with changing species composition both longitudinally and latitudinally, with no representative areas, and home to many species of mammals, birds, and other animals.

See Also the Following Articles

ARCTIC ECOSYSTEMS • FIRES, ECOLOGICAL EFFECTS OF • FOREST ECOLOGY • LATITUDE, COMMON TRENDS WITHIN • NORTH AMERICA, PATTERNS OF BIODIVERSITY IN • TIMBER INDUSTRY

Bibliography

Angelstam, Per K. (1998). Maintaining and restoring biodiversity in European boreal forests by developing natural disturbance regimes. *J. Veg. Sci.* 9, 593–602.

Bonan, G. B., and Shugart, H. H. (1989). Environmental factors and ecological processes in boreal forests. *Annu. Rev. Ecol. Syst.* 20, 1–28.

Boutin, S., Krebs, C. J., Boonstra, R., Dale, M. R. T., Hannon, S. J., Martin, K., Sinclair, A. R. E., Smith, J. N. M., Turkington, R., Blower, M., Byrom, A., Doyle, F. I., Doyle, C., Hik, D., Hofer, L., Hubbs, A., Karels, T., Murray, D. L., Nams, V., O'Donoghue, M., Rohner, C., and Schweiger, S. (1995). Population changes of the vertebrate community during a snowshoe hare cycle in Canada's boreal forest. *Oikos* 74, 69–80.

Cumming, S. G., Burton, P. J., and Klinkenberg, B. (1996). Boreal mixedwood forests may have no "representative" areas: Some implications for reserve design. *Ecography* **19**, 162–180.

Danell, K., Lundberg, P., and Niemelä, P. (1996). Species richness in mammalian herbivores: Patterns in the boreal zone. *Ecography* **19**, 404–409.

Elliott-Fisk, D. L. (1988). The boreal forest. In *North American Terrestrial Vegetation* (M. G. Barbour and W. D. Billings Eds.), pp. 33–62. Cambridge Univ. Press, Cambridge, UK.

Esseen, P.-A., Ehnström, B., Ericson, L., and Sjöberg, K. (1997). Boreal forests. *Ecol. Bull.* **46**, 16–47.

Hare, F. K., and Ritchie, J. C. (1972). The boreal bioclimates. *Geogr. Rev.* **62**, 333–365.

La Roi, G. H. (1967). Ecological studies in the boreal spruce–fir forests of the North American taiga. 1. Analysis of the vascular flora. *Ecol. Monogr.* **37**, 229–253.

Larsen, J. A. (1980). *The Boreal Ecosystem*. Academic Press, New York.

Peinado, M., Aguirre, J. L., and de la Cruz, M. (1998). A phytosociological survey of the boreal forest (*Vaccinio-Piceetea*) in North America. *Plant Ecol.* **137**, 151–202.

Qian, H., Klinka, K., and Kayahara, G. J. (1998). Longitudinal patterns of plant diversity in the North American boreal forest. *Plant Ecol.* **138**, 161–178.

Rowe, J. S. (1972). Forest regions of Canada, Publ. No. 1300. Canadian Forestry Service, Department of the Environment, Ottawa, Ontario.

Van Cleve, K., and Alexander, V. (1981). Nitrogen cycling in tundra and boreal ecosystems. *Ecol. Bull.* **33**, 375–404.

Walter, H. (1973). *Vegetation of the Earth in Relation to Climate and the Eco-Physiological Conditions* (J. Wieser Trans.). Springer-Verlag, Berlin.

BREEDING OF ANIMALS

David R. Notter
Virginia Polytechnic Institute and State University

GLOSSARY

breed "A homogenous, subspecific group of domestic livestock with definable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species, or . . . for which geographical separation from phenotypically similar groups has led to general acceptance of its separate identity" (Food and Agriculture Organization, 1995).

breeding line Any distinctive livestock population. The term may be synonymous with "breed" but is also often applied to somewhat distinct subpopulations within a breed or to breeding populations, either purebred or crossbred, that are of fairly recent origin. Breeding lines may exist within a breed, and newly created lines may become recognized breeds.

crossbreeding/crossbred Crossbreeding is the mating of animals of different breeds to produce commercial market animals or breeding animals. The resulting crossbred animals often exhibit improved fitness and performance compared to their purebred parents.

cryopreservation Long-term storage of gametes or embryos in liquid nitrogen.

***ex situ* preservation** Maintenance of a breed outside the environment and agricultural production system(s) in which it was developed and used. *Ex situ* preservation often involves cryopreserved gametes and/or embryos but may also involve live animals.

germplasm Sources of hereditary material. In animal breeding the term is commonly applied to breeding animals and to fresh or frozen sperm cells, ova, and embryos.

***in situ* preservation** Maintenance of a breed in the environment and agricultural production system(s) in which it was developed and used.

purebred An animal that is a member of a recognized livestock breed. Breed membership may be determined by pedigree records, geographical location, or knowledge of the breeding structure of the herd or flock. Animals that possess the full range of characteristics commonly associated with a specific breed are sometimes also designated as purebreds, but the designation must be recognized as subjective in these cases.

DOMESTIC ANIMALS make a critical contribution to human well-being as a source of food, fiber, power, and many other products and services. Worldwide, domestic animal production centers on a small number of globally distributed species, including cattle, pig, sheep, goat, chicken, duck, goose, and rabbit. These are augmented by a similar array of regionally important food- and fiber-producing species, such as buffalo, camel, llama, alpaca, yak, reindeer, turkey, and horse. There are several additional minor domesticated species, such as honeybee, guinea pig, and pigeon, and several species are currently undergoing domestication on a limited scale (e.g., various species of deer, antelope, game birds, ratites, and crocodilians and many aquacultural species). However, species-level domestic animal biodiversity is clearly very limited, representing *in toto* about 20 completely domesticated, and perhaps an additional 25 partially domesticated, species. Genetic diversity within domestic species, however, is vast, especially for the globally distributed species. Domestic animal genetic diversity evolved in synchrony with human dispersal as domestic animals accompanied their owners from their centers of domestication to every corner of the globe. Importantly, humans also provided these animals with the marginal levels of care, shelter, protection, and supplemental food that facilitated their adaptation to the diverse environments that they encountered. The resulting environmental adaptations, coupled with selection to meet a variety of human needs and preferences, resulted in the creation of an incredibly rich domestic animal fauna.

I. OVERVIEW OF LIVESTOCK GENETIC DIVERSITY

A. Breeds of Livestock

Populations of domestic livestock are traditionally subdivided and isolated from other subpopulations by both distance and management. These subpopulations are generally known as "breeds." A single, widely accepted definition of a livestock breed does not exist, but livestock breeds are approximately analogous to the subspecies or races of wild species. Thus, the Food and Agriculture Organization of the United Nations (FAO) defines a livestock breed as

> a homogenous, subspecific group of domestic livestock with definable external characteristics that enable it to be separated by visual appraisal from other similarly defined groups within the same species, or . . . for which geographical separation from phenotypically similar groups has led to general acceptance of its separate identity.

The concept of a livestock breed arose in nineteenth century England, where influential livestock breeders, including the noted Robert Bakewell, began to improve the native livestock of the region by controlled mating and selection of presumed superior individuals. Significantly, their activities included the establishment of "herd books" for most mammalian breeds which enumerated the pedigrees of animals within each breed so that breed membership was confirmed by records of ancestral ties to presumed elite foundation animals. These so-called "purebred" livestock thus had pedigree documentation of breed identity. Similar activities in poultry resulted in establishment of many avian breeds, although individual pedigree recording was not usually practiced for these smaller, more prolific species. This breed concept became widespread in Western Europe and in areas that were settled or influenced by Western Europeans. By the early twentieth century, a central part of livestock breeding in these areas was the identification and use of distinct, ostensibly superior, genetic types and the establishment of breeds to propagate these types.

The Western European definition of a breed had little meaning in most of Asia and Africa, where extensive pedigree recording did not appear to have arisen. However, selective breeding certainly did occur in these areas, and regionally distinct livestock types arose and were propagated. Isolation of these types was generally by distance or by selective matings based on obvious phenotypic traits such as color or horn shape rather than on pedigree information. However, pedigree information was occasionally accorded great value, as in the establishment of the Arabian horse. Today, the term breed has become globally pervasive and is used to describe any identifiable subpopulation possessing relatively distinct characteristics and maintained with a reasonable level of reproductive isolation.

B . Genetic Structure of Livestock Populations

Livestock breeds represent a partitioning of the genetic diversity within each species. Through many generations of selection and relative reproductive isolation, each of these subpopulations has become a distinctive genetic entity possessing specific combinations of genes and associated phenotypic characteristics. The core value of these breeds lies in the reliable access they provide to these combinations of genetically defined characteristics. For example, ewes of the prolific Russian Romanov sheep breed are capable of producing three to six lambs at a time. Sheep breeders interested in improving prolificacy can thus rely on this breed as a source of useful genes.

The traditional (i.e., mid-twentieth century) genetic structure of most livestock species was thus characterized by many regionally isolated subpopulations. In addition, several distinct subpopulations could often be identified within each region, maintained in more or less pure form by breeders who favored animals of a specific type. These dual mechanisms for maintaining reproductive isolation fueled development of levels of observed genetic diversity that are generally far in excess of those observed in wild species.

It would be incorrect, however, to view the livestock breeds as fixed, immutable genetic entities. Migration, or the exchange of genetic material, among breeds is not uncommon, and, even among breeds that maintain pedigree records, decisions to "open" the herd book occasionally occur to allow incorporation of desirable animals from outside the breed. Breeds wax and wane in popularity; as some breeds merge, others are absorbed by more popular breeds, and a few simply lose favor and gradually disappear. Breed evolution is thus an accepted part of livestock breeding. This traditional genetic structure of livestock breeds was remarkably similar to that proposed by Sewell Wright as optimal for evolution in natural populations, with relatively high levels of reproductive isolation, modest effective population sizes, diverse selection pressures, and levels of migration that provided for periodic infusions of genes without seriously compromising the genetic integrity of the subpopulations.

Livestock breeds in most cases possess very significant amounts of within-breed genetic diversity. Although breeds can rightly be viewed as mildly inbred lines, they do not approach the levels of genetic uniformity commonly found in plant breeding lines in which self-fertilization and clonal propagation can strongly limit, or in some cases completely eliminate, genetic diversity. Thus, essentially all livestock breeds retain significant evolutionary potential and can undergo significant genetic changes in response to selection, either natural or artificial. Livestock breeds retain the capacity to change, sometimes radically, in response to changes in breeders' preferences or market demands without losing the fundamental characteristics of the breed.

Finally, within each of the domestic species, a very large number of individuals exist that do not exhibit or possess unambiguous breed ancestry. In North American cattle, fewer than 10% of the animals are recorded as true purebreds. Worldwide, many domestic animals commonly display apparent breed affinity, manifested by the color or morphology of recognized local breeds, but lack either the pedigree documentation or the full array of phenotypic characteristics necessary to confirm breed identity. For example, more than 90% of U.S. dairy cows would be classified visually as belonging to the Holstein breed, but at most one-third of these would be recorded as purebred animals. The proportion of recorded animals is much lower in cattle used for meat production.

Many other animals are obvious crossbreds, showing mixtures of specific breed characteristics that suggest a somewhat predictable breed ancestry. Also, large numbers of animals are often truly nondescript, exhibiting a collage of various breed characteristics that point to a more catholic breed composition. The value of the crossbred or nondescript animals as a genetic resource is hotly debated. Populations are often large and well adapted to prevailing environmental conditions. Although adapted local breeds are often viewed as "contaminated" by crossing with less well-adapted, imported breeds, these populations also provide opportunity for creation of new genetic combinations and for selective elimination of undesirable breeds. Thus, these animals represent a potentially useful genetic resource but lack the predictability of the pure breeds for use in commercial livestock production.

C. Current Trends in Livestock Diversity

Perhaps the most significant event in global genetic resource utilization in the late twentieth century has been the emergence of highly productive, widely distributed "global" livestock breeds. These breeds, for the most part, arose in the developed nations of the temperate zones and have had a remarkable impact on livestock production worldwide. These breeds include the Holstein dairy cow, capable of producing an average of more than 50 kg of milk per day for 300 days; the white leghorn layer chicken, which can produce up to 300 eggs per year; the modern broiler chicken, which can reach a market weight of 2 kg in less than 40 days and on less than 4 kg of feed; meat-type pigs with less than 5 mm of back fat at a body weight of 110 kg; and Australian Merino sheep with wool fiber diameters of 17–20 μm.

The development and continued improvement of these global breeds is viewed in many quarters as the capstone of modern animal breeding. Large population sizes, detailed recording of performance, and intensive selection fueled their development. Well-organized breeders' groups and, recently, multinational corporations are involved in their propagation. The emergence of corporate animal breeding, in particular, allowed merging of genetic improvement, production, processing, and marketing activities and provided resources for aggressive international marketing of germplasm. Within the developed nations, the global breeds have made an important contribution to maintaining low food costs, but many of the traditional livestock breeds in these areas have declined because they cannot compete with the emerging global breeds.

Because of their tremendous production potential, these breeds also caught the attention of government planners and officials in developing nations. Their global dissemination was seen as a magic bullet to improve animal productivity, analogous to the improvements in crop yields associated with the Green Revolution. Widespread importations of these breeds into developing nations had a negative impact on indigenous breeds, sometimes through outright breed replacement but more often through extensive and generally unregulated crossbreeding between imported and indigenous breeds. In rural areas and subsistence production systems, the contributions of these highly productive global breeds were often disappointing. When denied the high levels of feeding, housing, and veterinary care under which they were developed and forced to grapple with unfamiliar levels of disease and climatic stress, the global breeds and their crosses were often unable to forage, survive, and reproduce at acceptable levels and were ultimately less productive than the adapted indigenous types they were intended to replace.

Improved technology for transfer and use of genetic material contributed to the demise of many indigenous breeds. Use of frozen semen became commonplace in cattle in the latter half of the twentieth century and is now widespread for many livestock species. Use of frozen embryos is also widely practiced for cattle, sheep, and goats. These developments allowed efficient global exchange of genetic material and effectively ended the genetic isolation of many breeds, especially in Asia and Africa, where the concept of breed identity was less strong than in Europe and North America. Greater ease of transport of live animals also facilitated establishment of populations of foreign breeds and replacement or contamination of native breeds, especially for the smaller or more fecund species such as poultry and pigs. Thus, modern reproductive technologies today permit foreign breeds to have a much greater and more rapid impact, and in some cases to completely replace native breeds.

Urbanization and economic development have also opened important niches for use of global breeds in developing nations. Demand for animal products in these countries has soared in recent years and is expected to continue to increase, resulting in movement from subsistence production to market-driven production and changes in production systems. In particular, urbanization has concentrated animal production in periurban areas to allow convenient access to population centers and led to opportunities for more intensive production. These changes have occurred despite concerns over economic disenfranchisement of farmers in rural areas, air and water pollution and waste disposal in periurban production, and the long-term sustainability of these production systems. Intensive animal production systems in periurban areas have also expanded use of high-production foreign breeds at the expense of adapted native types. The competitive advantage of these intensive periurban systems may decline once the full costs of waste disposal and environmental monitoring are incorporated into production costs, and a return to the more traditional mixed-farming systems and breed may be needed.

Expanded use of grain for animal feed would have been inconceivable to many international observers a decade ago but is now a reality. China has the highest aggregate consumption of meat in the world, and global grain production is projected to keep pace with demands for use in animal feed for at least the next 25 years. The result of these increases in use of grain for animal feed has been a need for rapid improvement in efficiency of grain use which has been difficult to achieve in a timely manner using traditional breeds. However, if trends in use of grain for animal feed slow or reverse in the latter years of the twenty-first century, these traditional breeds may again be needed.

Animal welfare concerns may also influence the use of livestock breeds. In Europe, concerns over issues of animal welfare in intensive production systems have prompted a return to less-intensive conditions and a corresponding need for animals that produce well under those conditions.

II. CONSERVATION PLANNING: COORDINATING THE USE AND PRESERVATION OF LIVESTOCK GENETIC RESOURCES

Animal breeders are often confronted by an apparent dilemma in the design of livestock breeding programs. On the one hand, maximizing livestock productivity relies on the identification and propagation of superior genetic types. Rates of genetic improvement are generally proportional to the intensity of selection. Thus, propagation of small numbers of elite parents yields greater rates of progress than retention of larger numbers of less desirable individuals. Indeed, most of the accomplishments of modern animal breeding have resulted from improving the accuracy of genetic evaluation of prospective parents and increasing the impact of these selected parents by use of artificial insemination, embryo cloning and transfer, and other advanced reproductive technologies.

Successes of modern animal breeding programs have, in some cases, been spectacular, as exemplified by the emerging global breeds of poultry, swine, and dairy cattle. The global proliferation of industrial production systems has relied primarily on these species because of their greater efficiency of use of harvested feeds. In other species, such as the grazing ruminants, successes have been more modest, largely because these animals are generally produced under extensive conditions and required to interact closely with their environment to harvest forages, but still have been substantial. Despite these successes, demands for further increases in rates of animal production to supply an expanding human population continue to increase, and pressure to utilize the highly productive global breeds to meet these demands is tremendous.

Improvement-oriented animal breeders view global genetic diversity as a resource to be used in pursuit of improved animal productivity. The existing array of livestock breeds is recognized as potentially valuable but primarily as a consumable resource to be integrated into elite populations as appropriate and molded into ever-improving commercial populations. This viewpoint, described as the “utilizationist view” by the U.S. National Academy of Science, emphasizes the potential value of the unique genes and gene combinations found in the various breeds as raw material for breed evolution and emphasizes the value in maximizing that evolution. In the utilizationist view, there may well be a need for a few dozen different pig breeds to meet the demands of different global production environments and markets, but a commitment to maintain all of the 353 currently recorded pig breeds would be seen as excessive and counterproductive.

In contrast to this position is a more “preservationist view” which tends to accord individual breeds a position similar to that of an endangered species and proposes similar management strategies to sustain them. The preservationist view emphasizes the unique history and presumed genetic distinctness of individual breeds and takes breed preservation per se as its goal. Blending of breeds to generate improved, adapted commercial populations is acknowledged to be necessary and desirable, but only against a backdrop of secure and relatively stable populations of the contributing breeds. Retention and continued use of existing breeds in their traditional environments and production systems is promoted, even if public subsidies are required to ensure that use.

Synthesis and rectification of the utilizationist and preservationist views are badly needed but have yet to fully occur in the animal breeding community. The utilizationists have, to some extent, become victims of their successes, especially in poultry, dairy cattle, and pig production. Aggressive and widespread sampling and comparative evaluation of breeds in the mid-twentieth century led to the establishment of today’s global breeds and to the replacement of many of the middle-level breeds that were widely represented in commercial production a few decades ago. Although large numbers of locally adapted, relatively lowly productive breeds of poultry and swine still exist globally, the elite global breeds have become so differentiated from these stocks that there is now thought to be minimal opportunity for these breeds to contribute genetic material to the elite global breeds through traditional breeding methods. Thus, industrial stocks of chickens and turkey have experienced no outside contributions from other breeds in more than 25 years. A similar situation appears to be evolving in modern pig populations, and, at least in temperate regions, the Holstein dairy cow is rapidly becoming the preeminent, and in many cases exclusive, dairy breed. Increasingly, the elite global breeds are viewed as dependent on existing reservoirs of intrabreed genetic diversity for future adaptation and improvement.

In contrast, the preservationist position has been made untenable by the events of global economic and social homogenization. In an increasingly cosmopolitan, interdependent, and rapidly changing world, continued use of 5000 livestock breeds will not occur and probably is not necessary to maintain adequate reservoirs of genetic diversity. However, a responsible fall-back position which can ensure the identification, retention, and ready access to the core genetic diversity of each domestic species is badly needed.

A reasonable synthesis and integration of conservation, preservation, and improvement activities have been achieved in the plant breeding community, as described elsewhere in this encyclopedia. Facilities for long-term seed storage exist and have both a broad mandate and the budgetary support to acquire and preserve samples representing the full range of diversity within individual species of crop plants and their wild relatives (Box 1). The importance of these programs is acknowledged by both public agencies and private breeding companies. Although there are fundamental differences between plants and animals in the population structure of breeding materials and in the technologies that can be used for genetic resource preservation and use, a more proactive approach to animal germplasm conservation is needed.

Box 1

Use of Wild Relatives in Livestock Breeding

In contrast to plants, use of wild relatives as a source of genetic material for industry breeding programs in livestock is limited. In most cases, domestic animals represent distinctive species and wild progenitors of these species are no longer found. However, several recent developments suggest that wild relatives may be a useful source of genetic material and that the pool of useful wild relatives may be broader than originally thought.

In bovines, hybrids of domestic cattle with gaur (*Bos gaurus*), banteng (*B. javanicus*), and yak (*B. mutus*) are common in Asia and make important contributions to livestock production. In North America, hybrids of cattle and bison (*Bison bison*) were evaluated in Canada in the 1960s but with limited success because of near-complete infertility of hybrid males. Recent work in the United States, however, yielded enough fertile hybrid males to allow establishment of herds of animals, known as beefalo, possessing a mixture of genes from cattle and bison. These animals were claimed to possess some desirable adaptational and meat quality traits but are today primarily a novelty.

In western China, genes from the wild ibex (*Capra ibex*) have been introduced into domestic cashmere goats. The yield of cashmere from ibex is very low, but the cashmere fibers are very fine, averaging about 12 μm in diameter. Typical cashmere goat fibers are 16 or 17 μm in diameter, and the value of the fiber is related to its fineness. Introgression of genetic material from ibex, followed by selection for yield and fineness of cashmere, led to development of animals with 12.5–25% ibex genes that possess increased fineness of the cashmere with near-normal yield. Similar crossing programs have been proposed using the wild vicuña to improve fiber quality in domestic alpaca.

In the Republic of South Africa, there is a need for hardy and well-adapted livestock for use in communal and other low-input production systems. In pigs, crosses between an endangered domestic pig breed, the Kolbroek, and the wild bush pig (*Potamochoerus porcus*) led to production of fertile offspring with improved hardiness, foraging ability, and capacity to digest fiber. This intergeneric cross also indicates the potential to use more distant crosses as a source of genetic material.

Wild relatives of domestic livestock have also contributed to development of genetic maps. Crosses of domestic pig with wild boar (*Sus scrofa*) were used to develop gene maps for pig, and crosses with red jungle fowl (*Gallus gallus*) were used in mapping the chicken genome.

In plant breeding, wild relatives of domestic crop plants have long been recognized as a source of genes for disease and pest resistance. Recent work with wild relatives of tomato and rice suggest similar opportunities to improve production traits. Use of gene maps to identify useful genetic material and of improved methods to introduce this material into domestic stocks using either biotechnology or conventional breeding methods now exist for plants and will be developed for animals.

III. MANAGEMENT STRATEGIES FOR CONSERVATION OF LIVESTOCK GENETIC RESOURCES

Leadership in conservation of livestock genetic resources at the global level has come from the Food and Agriculture Organization of the United Nations (FAO), beginning with many *Animal Production and Health Papers* and continuing with the current Domestic Animal Diversity Information System (accessible via the worldwide web at *http:www.fao.org/dad-is*). In particular, the FAO *Global Strategy for the Management of Farm Animal Genetic Resources,* and the associated *Guidelines for Development of National Farm Animal Genetic Resource Management Plans,* provide for international communication and cooperation in managing domestic animal diversity (Box 2).

A. Inventory, Description, Characterization, and Assessment of Degree of Endangerment

An understanding of the status of domestic animal genetic diversity is a prerequisite to its successful management, and the inventory, description, and comparative characterization of livestock breeds is a key management activity. The FAO *World Watch List for Domestic Animal Diversity,* currently in its second edition, summarizes the contents of the FAO Global Databank, with

Box 2

Crossbreeding: Threat or Boon to Animal Genetic Resources?

The production of crossbred animals by mating parents of different breeds is common in livestock breeding. Crossbred animals are often superior to their purebred parents and therefore more desirable for commercial farmers. The desirability of crossbred livestock arises from two sources: heterosis and complementarity.

Heterosis, also known as hybrid vigor, is normally manifested as an increase in fitness of crossbreds relative to their purebred parents. Thus, crossbred animals commonly are more fertile, more disease resistant, and better able to cope with environmental stresses than would be predicted from the average fitness of their purebred parents. As a result, productivity traits such as growth and milk or egg production are also usually improved in crossbred animals. In simple terms, purebred individuals commonly show modest levels of inbreeding as a result of the restricted matings required to genetically fix the defining characteristics of the breeds (e.g., colors and horn shape), and inbreeding commonly results in associated modest reductions in fitness. Crossing of purebred animals relieves accumulated inbreeding and provides an economically significant "kick" in performance. Significantly, however, effective use of hybrid vigor requires the crossing of unrelated parent breeds and is maximally expressed only in the first-generation cross. Thus, hybrid vigor can be maximally exploited and reliably captured only when parents of the original breeds are maintained as purebreds and crossed in very specific ways.

Complementarity is a characteristic of the production system which arises when different breeds play different and appropriate roles in crossbreeding systems. In low-input or extensive production systems, adaptation of breeding females to the production environment is critical. If the environment is harsh, the productive capacity of indigenous breeds is often low, thereby allowing demands for nutrients and other inputs to be synchronized with their limited supply. In meat production, mating of females of indigenous breeds to males of more productive and heavily muscled breeds can increase the value of the offspring while maintaining high levels of adaptation in the breeding females. Additionally, the benefits of having a well-adapted mother and of hybrid vigor often permit the crossbred offspring to perform at acceptable levels and increase overall productivity. However, successful use of complementarity again requires maintenance of the adapted, indigenous breeds. Replacement of the original indigenous breeding females with crossbreds results in losses in both adaptation and hybrid vigor, often with serious negative effects on productivity.

Proper use of crossbreeding thus relies on breed diversity. Access to arrays of both adapted and highly productive breeds allows synchronization of the genetic characteristics of both breeding females and crossbred offspring to diverse production environments and markets. However, proper use of crossbreeding also requires that breeds be maintained in their proper roles in the system. In developing nations, in which control over matings is often limited or nonexistent, use of highly structured crossbreeding systems has proven difficult, and initial improvements in productivity in first-generation crosses, as well as the genetic integrity of the indigenous breeds, have been lost through indiscriminant mongrelization of adapted and unadapted breeds. Thus, crossbreeding has significant benefits when an appropriate management infrastructure exists but can have devastating effects on indigenous breeds when not adequately managed.

emphasis on breeds that are at risk because of limited population size, uncontrolled crossbreeding, or declining use.

The FAO Global Databank lists 3019 breeds representing 14 domestic species of mammals and 863 breeds representing 11 avian species (Tables I and II). Of these, data on population size and status are available for 2191 mammalian and 733 avian breeds, whereas the remaining breeds possess only cursory information on breed characteristics and distribution. Reporting of breed information is not consistent throughout the world. European breeds dominate the databank with 1172 mammalian and 516 avian entries, primarily because of a long history of concern over genetic diversity in Europe and because Europe was the cradle for the concept of breed formation. African breeds, in contrast, are almost certainly underreported, with only 291 mammalian and 105 avian breeds listed. The total number

TABLE I

Global Summary: Mammalian Breeds at Risk, by Species

Species	No. of breeds				
	On file	With population data	Endangered[a]	Critical[a]	At risk[b]
Buffalo	72	55	2	0	2
Cattle	787	582	77	58	135
Yak	6	6	0	0	0
Goat	351	267	31	13	44
Sheep	920	656	85	34	119
Pig	353	265	39	30	69
Ass	77	24	2	7	9
Horse	384	277	87	33	120
Dromedary	50	40	1	1	4
Bactrian camel	7	7	0	1	2
Alpaca	4	4	0	0	0
Llama	3	3	0	0	0
Guanaco	2	2	0	0	0
Vicuña	3	3	0	0	0
Total	3019	2191	324	177	501

[a] See Table III for definitions.
[b] Sum of endangered and critical breeds taken from FAO (1995).

of mammalian livestock breeds is thus projected to be about 4000 and the total number of avian breeds likely exceeds 1000. To place these numbers in context, the number of breeds for these 14 domestic species of mammals approximately equals the number of mammalian species on Earth.

To be useful, inventory information must be augmented with some objective assessment of degree of

TABLE II

Global Summary: Avian Breeds at Risk, by Species

Species	No. of breed				
	On file	With population data	Endangered[a]	Critical[a]	At risk[b]
Chicken	606	512	227	47	274
Domestic duck	62	54	18	11	29
Turkey	31	29	6	5	11
Muscovy duck	14	13	4	1	5
Domestic goose	59	51	18	10	28
Guinea fowl	22	17	3	1	4
Partridge	11	4	0	0	0
Pheasant	8	7	0	0	0
Quail	24	23	0	16	16
Pigeon	19	16	0	2	4
Ostrich	7	7	0	3	3
Total	863	733	276	96	372

[a] See Table III for definitions.
[b] Sum of endangered and critical breeds taken from FAO (1995).

TABLE III

Criteria for Determining Degree of Endangerment for Breeds of Domestic Livestock[a]

Category	Criteria
Not at risk	Total number of breeding females is greater than 1000 and total number of breeding males is greater than 20, or Population size approaches 1000, the percentage of females being bred to males of the same breed is near 100%, and the overall population size is increasing
Endangered	Total number of breeding females is between 100 and 1000 or the total number of breeding males is between 5 and 20, or Overall population size is slightly below 100 and is increasing and the percentage of females being bred to males of the same breed is more than 80%, or Overall Population size is slightly more than 1000 but is decreasing and the percentage of females being bred to males of the same breed is less than 80%
Critical	Total number of breeding females is less than 100 or the total number of breeding males is less than 5, or Overall population size is slightly more than 100 but is decreasing and the percentage of females being bred to males of the same breed is less than 80%
Extinct	It is no longer possible to recreate the breed population from live animals and/or cryopreserved gametes or embryos
Unknown	Information on population size is not available

endangerment of the populations involved. The FAO guidelines for determining breeds at risk are shown in Table III. Based on these guidelines, about 23% of mammalian and 51% of avian breeds with population data are considered at risk. Furthermore, only 27% of mammalian breeds and 48% of avian breeds that are at risk have active programs in place to promote their conservation. The combined total of 30% of breeds at risk, if extrapolated to the projected 5000 global breeds, results in a figure of about 1500 breeds that are vulnerable to loss or to serious erosion of genetic diversity. Because many of these endangered breeds are found in harsh production environments with minimal care and management, they can reasonably be anticipated to possess unique genetic adaptations and disease-resistance characteristics. However, in common with many endangered species, they are at risk of disappearing before they can be adequately evaluated and characterized.

B. Preservation Strategies

Conservation of livestock genetic resources will almost certainly involve both *in situ* preservation of breeding herds in agricultural production systems and *ex situ* preservation of animals, gametes, and embryos in reserves and cryogenic repositories. Issues of current utility, degree of endangerment, and cost of preservation will determine the method(s) of choice.

1. *In Situ* Strategies

Strategies that are based on effective use of livestock genetic diversity in commercial and/or subsistence production systems are most likely to provide the security necessary to allow breeds to remain viable and to maintain the population sizes necessary to ensure continued improvement. Thus, a key activity in livestock genetic resource conservation is the comprehensive appraisal of the productive merit of the various breeds under current productive conditions.

Most of the world's domestic animals are found in the developing nations, and nearly 2 billion people obtain at least part of their daily livelihood from livestock. Rapid changes in production conditions have occurred in some of these nations, but the number of animals that are used in subsistence production systems remains very large. Levels of veterinary care, supplemental feeding, and shelter from climatic stresses often remain limited, and adaptational characteristics of indigenous livestock remain important to many farmers. Studies throughout the world have confirmed that when environmental conditions are harsh, imported temperate breeds of high production potential often experience increased levels of mortality and morbidity and reduced reproductive success, and as a result they may be less productive per unit time or per unit of inputs than indigenous breeds. Often, advisers recommend changes in the production environment, even though farmers often do not have the financial resources or market incentives to do so. A more responsible strategy, and certainly one that is more consistent with conservation of livestock diversity, is to focus on the retention and improvement of local breeds. The implementation of modern livestock improvement procedures in indigenous breeds may permit them to be improved in parallel with gradual improvements in production conditions while retaining important adaptational characteristics.

An example of such an approach is found in India, where Operation Flood was begun more than 30 years ago to enhance milk production and marketing opportunities for smallholders in western India. The goal was to improve the nutrition of the people of India by increasing supplies of buffalo and cattle milk. Certainly, Operation Flood and its successor projects have been a quantitative success. India is now the world's number 1 producer of milk. More important, this success was achieved without recourse to establishment of high-volume Western dairy production systems or a rapid infusion of poorly adapted foreign germplasm. Instead, emphasis was placed on establishing marketing conditions that would reward efficient producers and provide incentives for improved production practices. Only recently has the identification of superior germplasm within the indigenous breeds of cattle and buffalo and the strategic use of imported breeds become a part of this program.

In other cases, however, the pace of social change has been so rapid that indigenous breeds in the developing nations cannot meet the challenges. Under these conditions, pressures on indigenous breeds can become extreme very quickly, and careful analysis is required to identify appropriate genetic resource development activities. Developments in pig production in China provide an example. China has the world's most diverse array of pig breeds, with at least 50 and perhaps as many as 100 distinct breeds. Almost all evolved in situations in which grain was primarily used for human food. The pigs therefore became adapted to the use of fibrous by-product feeds from grain and vegetable production. They also were relatively slow growing (to synchronize their nutrient requirements to their limited food supply), quite fat (because animal fat was an important energy source for peasant farmers), and, in some cases, highly prolific. In the 1980s, government policies in China were changed to provide greater access to grain for animal feed. Under these conditions, the indigenous pig breeds were markedly inferior to the faster growing, leaner, and more feed-efficient Western breeds. Government-sponsored importations and entry of multinational pig breeding companies into the market resulted in widespread crossbreeding of indigenous pigs with Western breeds. An extensive network of government artificial insemination (AI) centers for pigs facilitated the use of imported breeds, and the development of more sophisticated urban markets heightened preferences for leaner pork. By the early 1990s, many of the AI stations no longer provided boars of the local breeds.

The events in Chinese pig breeding were fully rational responses to changed government policy and to changes in the economy and the nature of society. Both farmers and consumers benefited from the changes. However, the result has been to endanger a whole array of local breeds, many of which possess globally unique characteristics.

In developed nations, focus on a declining number of elite breeds in relatively intensive production systems has likewise endangered many traditional breeds. The situation is particularly acute in Europe, and greater cognizance of the problem exists there. The European Association of Animal Production Working Group on Animal Genetic Resources in 1993 listed 877 European breeds of cattle, sheep, goats, pigs, and horses. Of these, less than half (412 breeds) were considered secure. The remainder were classified as "potentially endangered" (162 breeds), "minimally endangered" (89 breeds), "endangered" (43 breeds), or "critically endangered" (158 breeds). No information was available for 13 breeds.

In response to this imminent contraction in livestock genetic diversity, many nations of Western Europe have initiated programs to maintain endangered breeds. Although often using a combination of *in situ* and *ex situ* techniques, high priority is placed on *in situ* preservation. The importance of livestock breeds as a cultural and historical resource reflecting the heritage of the nation or as a component of unique and perhaps themselves endangered agroecosystems is particularly recognized in Western Europe. The contribution of livestock breeds to public welfare is sometimes referred to as "landscape value," and public resources are increasingly being directed toward breed conservation activities. These programs may involve maintenance of endangered breeds on public facilities or direct payments to farmers who maintain endangered breeds.

Throughout the developed nations, "grassroots" organizations play a significant role in *in vivo* preservation of endangered livestock breeds. These nongovernmental organizations (NGOs) have for many years provided leadership for conservation of livestock diversity. Examples include the Rare Breeds Survival Trust in the United Kingdom, the American Livestock Breeds Conservatory in the United States, and the Canadian Foundation for the Conservation of Farm Animal Genetic Resources. The success of these organizations attests to the potential to involve committed private individuals in breed conservation. They have been most successful in wealthy countries in which significant private resources can be exploited, but the identification of private patrons to aid in breed conservation can also occur in developing nations. The success of these organizations relies on members who are committed to, and educated about, the technical aspects of genetic conservation. Otherwise, well-meaning but misdirected breeding policies may result in losses of genetic diversity. Public–private partnerships to provide members of these NGOs with technical expertise can be particularly beneficial.

2. *Ex Situ* Strategies

Ex situ preservation usually involves cryopreservation of gametes and embryos but may also involve live animals kept in farm parks, on research farms, or in other noncommercial settings. In the developed nations, farm parks have increased in popularity and often contribute to meaningful maintenance of biodiversity through their association with responsible grassroots organizations. Live-animal, *ex situ* preservation can also occur at publicly funded facilities, generally under conditions that at least approximate those found in commercial agriculture but which often do not replicate the particular conditions under which the breeds evolved and were traditionally used. In these situations, selection for unique adaptational characteristics is relaxed but with proper breeding management and adequate population size, the key genetic characteristics of the breed can be retained for many generations. Costs of *ex situ* live-animal preservation are high, however, involving feeding and daily care of breeding animals. The risks associated with live-animal *ex situ* programs are therefore also high, especially in developing nations in which sustained funding for long-term conservation programs may not be available and in which risks of social and political upheaval are greatest.

The use of cryopreserved gametes, embryos, and tissues is a more common form of *ex situ* preservation. In farm animals, sperm cells can be successfully frozen and stored for future use in all species, although success rates from use of frozen sperm vary considerably among species. Techniques for collection, cryopreservation, and subsequent use of sperm cells are relatively well developed, a single collection provides a relatively large number of gametes, and multiple samples are relatively easy to obtain. For these reasons, cryopreserved sperm cells are the most common material used for *ex situ* preservation of endangered breeds.

Cryopreserved sperm are ideally suited to support *in situ* preservation activities. Storage of sperm from a wide sample of males of a breed provides future access to the genetic material of these representative foundation animals. Losses of genetic diversity in living populations can thus, if necessary, be restored by use of sperm from males of past generations.

However, cryopreserved sperm cells are not particularly efficient for regeneration of a breed that has become extinct. Sperm cells contain only a sample half of the animal's DNA; therefore, restoration of an extinct breed from cryopreserved sperm requires a "grading-up" process in which sperm is used on females of a different breed over several generations to eventually create animals that have a majority of their nuclear genes from the cryopreserved breed. The efficiency of use of sperm cells to restore a breed depends on the generation time and the fecundity of the species. For example, restoration of 93.75% of the nuclear genes (i.e., four generations of upgrading) can be accomplished in about 3.5 years in pigs, and use of 100 sows initially could conservatively result in production of 1000 breeding females by generation 4. In contrast, in cattle a minimum of 10 years would be required using conventional breeding techniques to produce even a small number of animals possessing 93.75% of the nuclear genes of the preserved breed.

Also, although nuclear genes can be adequately preserved using frozen sperm, cytoplasmic DNA found in animal mitochondria are contributed only via the ovum and can thus be preserved only by storage of cryopreserved ova or embryos. Cryopreservation and subsequent *in vitro* fertilization of ova are not yet practical for domestic species. Embryo cryopreservation, however, is practical in cattle, sheep, and goat and permits preservation of the full genome, both nuclear and cytoplasmic. However, collection of embryos for cryopreservation is more difficult and expensive than collection of sperm cells and yields of embryos are much lower, often on the order of only two to eight embryos per mating. Also, embryos of pig and poultry cannot be reliably cryopreserved.

Current recommendations for *ex situ* preservation programs thus generally focus on extensive use of frozen sperm cells. However, small numbers of cryopreserved embryos (in species for which this is possible) and/or small populations of breeding females preserved under either *in situ* or *ex situ* conditions generally must also be maintained to serve as a source of cytoplasmic genes and to allow efficient regeneration of the breed. A store of cryopreserved embryos produced by mating 25 pairs of unrelated parents and bolstered by cryopreserved sperm from 25 unrelated males can effectively capture the genetic diversity of a breed for long-term preservation and future use.

Cryopreservation of somatic cells is also becoming common in *ex situ* preservation programs under the assumption that cloning of new individuals from cryopreserved adult cells will eventually become a reality. Use of cloning to support preservation of rare breeds will be discussed in Section IV,C.

IV. POTENTIAL IMPACTS OF NEW TECHNOLOGIES

A. Advances in Cryopreservation

Several technological advances in cryopreservation promise to enhance our ability to preserve biodiversity. Key research areas in cryopreservation include

1. Cryopreservation of pig embryos to allow capture of the full genome including mitochondrial DNA. This is especially important in Asia, where the number of endangered pig breeds is large.
2. Cryopreservation of poultry embryos and semen. In chickens, success rates from use of cryopreserved semen remain low. Better success rates would prompt greater use of frozen semen in the poultry industry, which would then facilitate storage and access to frozen gametes for conservation. Cryopreservation of fertilized eggs is likely to remain difficult. However, it appears possible to extract and cryopreserve the germinal disk of cells from the fertilized egg separate from the egg itself.
3. An improved understanding of why semen from some males and embryos from some breeds do not survive freezing well. Large individual differences exist in success rates for cryopreservation of semen and embryos. In commercial cattle breeding, there is considerable selection for "freezability" of semen, and because the populations are large this selection has little impact on the genetic diversity within the breed. However, endangered breeds are commonly represented by only a few males, and if semen from several of these cannot be satisfactorily cryopreserved, a significant segment of the biodiversity of the breed may be lost. Success rates for frozen embryos are also often lower in poorly characterized endangered breeds, and these breeds sometimes respond poorly to the hormonal regimen required for embryo collection. More reliable and robust techniques for semen and embryo recovery and cryopreservation are thus needed.

B. Gene Mapping

The development of comprehensive gene maps is occurring rapidly for several domestic species. These maps will facilitate screening of breeding animals and of cryopreserved material for unique genetic variants. Understanding of the molecular genetic control of quantitative traits is still in its infancy but is expanding rapidly. Identification of quantitative trait loci (i.e., regions of the DNA which influence performance) is beginning and promises to aid in the screening of rare breeds for useful genes. Improved gene maps will also aid in determining genetic relationships among breeds, thereby assisting in prioritization of candidates for preservation. Long-term conservation of the 5000 or more global breeds of mammalian and avian breeds is unlikely to occur, but knowledge of relationships among breeds would allow identification of a smaller "core collection" of breeds representing the bulk of the genetic diversity within each species.

C. Cloning

Cloning of farm animals from adult cells, if perfected and commercialized, can have far-reaching effects on both the use of biodiversity and the efficiency of its preservation. Certainly, the widespread use of genetically identical cloned individuals in commercial production systems would reduce the biodiversity present within those systems and increase the genetic vulnerability of the animals to specific diseases or other environmental stressors. Preliminary risk assessment studies are already underway to attempt to optimize the number of clonal lines that should be used within a herd or flock to minimize these dangers.

The ability to efficiently produce clonal offspring from adult animals (i.e., from individuals whose performance has already been characterized), however, has potential to facilitate synchronization of genetic resources to specific production conditions. In such a scenario, breeding animals that have demonstrated exceptional adaptation and productivity in a particular environment could be clonally reproduced for use in that specific environment.

The ability to produce clonal offspring from adult cells would have tremendous implications for preservation of endangered genetic resources. Large numbers of cells could be harvested from individual animals using minimally invasive biopsy techniques and cryopreserved for future use. Production of clonal offspring from these cryopreserved cells would recreate the full genetic complement (both nuclear and cytoplasmic) of the preserved animals. Furthermore, because sampling of somatic cells for cloning is relatively easy, large numbers of founder animals could be sampled, increasing the biodiversity present within the sample. Complications involved in harvesting and preservation of gametes and embryos would be circumvented and efficiency of storage of animal genetic material would approach that enjoyed in plants by storage of seeds.

Although technology for clonal propagation of adult animals cannot yet be relied on for preservation of animal genetic resources, preliminary results hold great promise. Dolly, the sheep produced by cloning of adult mammary cells at the Roslin Institute in Scotland, represented a watershed event in this technology. Subsequent cloning of cattle and pigs from adult cells has occurred, and recent reports in the popular press (eagerly awaited in the scientific literature) indicate that easily accessible bovine cells obtained from the ear may be satisfactory for cloning. Already, one of the last individuals of an endangered New Zealand breed, the Enderby Island cow, has been cloned to attempt to save the breed.

V. CONCLUSIONS

An assessment of likely trends in global livestock genetic diversity must recognize that demands of population growth, economic development, and urbanization will require significant increases in animal productivity worldwide. The normal processes of breed evolution have clearly accelerated during the past 25 years. Breeds are disappearing at more rapid rates than in the past and the rate of new breed formation has not correspondingly increased. Many indigenous breeds are ill-suited to contribute to meeting those challenges in a timely manner and will likely become endangered or restricted to subsistence production systems. Global, temperate breeds, with their high production potentials, will contribute to world food production, placing further pressure on less-productive indigenous breeds. However, the adaptational characteristics of local breeds have great value in helping to synchronize animal genetic resources with local stressors and should contribute to development of new, improved breeds. Thus, a new synthesis of livestock genetic resources is needed in which genetic resources from a variety of sources are managed to maximize livestock productivity in sustainable production systems, and this synthesis must be accompanied by programs to protect breeds which may be endangered or replaced.

See Also the Following Articles

BREEDING OF PLANTS • CAPTIVE BREEDING AND REINTRODUCTION • CATTLE, SHEEP, AND GOATS, ECOLOGICAL ROLE OF • GENETIC DIVERSITY • INBREEDING AND OUTBREEDING

Bibliography

Council for Agricultural Science and Technology (CAST) (1999a). Animal agriculture and global food supply, Task Force Report No. 135. CAST, Ames, IA.

Council for Agricultural Science and Technology (CAST) (1999b). Benefits of biodiversity, Task Force Report No. 133. CAST, Ames, IA.

Food and Agriculture Organization (FAO) (1995). *World Watch List for Domestic Animal Diversity*, 2nd ed. FAO, Rome.

Food and Agriculture Organization (FAO) (1998a). *Management of Small Populations at Risk*. FAO, Rome.

Food and Agriculture Organization (FAO) (1998b). *Primary Guidelines for Development of National Farm Animal Genetic Resource Management Plans*. FAO, Rome.

International Food Policy Research Institute (IFPRI) (1999). Livestock to 2020: The next food revolution, Food, Agriculture, and the Environment Discussion Paper No. 28. IFPRI, Washington, DC.

National Research Council (1993). *Managing Global Genetic Resources: Livestock*. National Academy Press, Washington, DC.

Oldenbroek, J. K. (Ed.) (1999). *Genebanks and the Conservation of Farm Animal Genetic Resources*. DLO Institute for Animal Science and Health, Lelystad, The Netherlands.

Simon, D. L. and Buchenauer, D. (1993). *Genetic Diversity of European Livestock Breeds*, European Association of Animal Production Publ. No. 66. Wageningen Press, Wageningen, The Netherlands.

BREEDING OF PLANTS

Donald N. Duvick
Iowa State University

GLOSSARY

biodiversity (biological diversity) Hereditary variation in life-forms at all levels of organization. Examples are diversity among wild species, among individuals within those species, among crop species, and among and within varieties of a given crop.

crop varieties Named populations of crop plants that possess recognizable features and known utility for food, feed, or fiber.

durable resistance Inherited resistance to a disease or insect pest, with relatively long effective lifetime. It is usually imparted by several genes of individually small effect.

genotype Genetic constitution of an organism or group of organisms such as a plant or plant variety.

heterosis Hybrid vigor, exhibited when offspring are larger, more vigorous, and more productive than the parents.

monoculture Usually refers to growing a single uniform plant variety over a large area.

multilines Planned mixtures of different selections of the same variety that differ only in genes for disease or insect resistance.

polyculture Usually refers to growing several crops in mixture or in rotation, often with variability among the plants of a given crop population.

specific resistance Narrowly targeted resistance to a specific genotype of disease or insect pest. It is usually imparted by a single gene of large effect, and typically with a relatively short effective lifetime.

traits Recognizable characteristics of a crop variety, such as plant height, grain color, specific disease resistance, yield potential, or tolerance to heat and drought.

transgenes Genes imparted to an organism by means of biotechnology rather than sexual hybridization.

THE GOAL OF CROP AGRICULTURE IS TO REDUCE BIODIVERSITY in favor of increased and easily accessible food supplies. This was true for the first farmers and it is still true today. The reduced biodiversity brings advantages but also disadvantages to farmers—to food producers—because biodiversity is essential for the maintenance of ecosystem stability. The following discussion shows how farmers and plant breeders have dealt with this paradox through the millennia up to the present time, and how they can cope with it in the years to come.

Box 1

E. O. Wilson defines biodiversity (biological diversity) as “the totality of hereditary variation in life forms, across all levels of biological organization, from genes and chromosomes within individual species to the array of species themselves and finally, at the highest level, the living communities of ecosystems such as forests and lakes” (Wilson, 1994).

Box 2

The biodiversity found among and within crop varieties is a subsection of the totality described by Wilson. Today’s professional plant breeders usually describe that subsection as “genetic diversity.”

I. THE FIRST FARMERS

Ten thousand years ago in western Asia, and later in eastern Asia and the Americas, hunter-gatherers began to specialize in gathering and nurturing a few favorite species of plants (Smith, 1995). Eventually the favored species were purposely planted in isolated plots, seed from the most desirable plants was replanted, and gradually domesticated crops were invented.

The domesticates represented only a few of the hundreds of species that had furnished food for the hunter-gatherers. The fields in which they were grown were cultivated and weeded, and the farmers did their best to eliminate all plants except the ones they favored. They depended not only on a greatly reduced number of species for the majority of their food supplies, but they also used only a portion of the total genetic variability within each species. They selected only those rare plants in a species that had the traits needed for domestication, such as nonshattering seeds or uniform germination. As a result, plant biodiversity in food production was greatly reduced.

But within each of the newly domesticated crops, farmers selected and reselected for different kinds of desirable plant and seed types, gradually increasing the diversity in the narrow base that had been provided by the original selections. Within a species, groups of related individuals were selected for special grain types, for adaptation to particular heat and drought conditions, or for resistance to disease or insect pests. In this way the farmers developed what we now call crop varieties. The varieties were recognized individually for their unique constellations of useful traits, and often they were given names that described their most important features. Biological diversity, in this case, genetic diversity, had been increased within each crop species. This diversity came in part from recombination of traits in the original selections, in part from genetic mutation, and in part from accidental backcrossing to the still-existing wild species in those cases where they still grew in proximity to the farmers’ fields.

Although the reduction of biodiversity in favor of carefully tended plantings of specially selected varieties increased amounts and availability of food from favored species, it brought its dangers as well. Uniform and broad expanses of one crop enabled easy multiplication of diseases and insects especially adapted to the particular genetic constitution of the farmers’ varieties. Catastrophic attacks of disease, or invasions of insects, or unprecedented heat and drought, or unusually cold and wet growing seasons caused wholesale crop destruction and ensuing famines whenever the farmers’ crops had insufficient diversity to provide at least some plants or some species with ability to withstand the assaults. These problems, a consequence of lack of the right kinds of genetic diversity among and within crop varieties, are as old as agriculture. They are recorded in myth and in written history, and they still occur wherever crop agriculture is practiced. Plant breeders today use the term “genetic vulnerability” to describe the condition that results “when a crop is uniformly susceptible to a pest, pathogen, or environmental hazard as a result of its genetic constitution, thereby creating a condition for disaster” (Board on Agriculture and National Research Council, 1993).

The phrase “genetic vulnerability” is new but the condition is not. Red rust on wheat (*Triticum* spp.) in Roman times, mass poisoning from ergot-tainted rye (*Secale cereale* L.) in medieval Europe, potato late blight and the Irish potato (*Solanum tuberosum* L.) famine of the nineteenth century, and the widespread Southern Corn Leaf Blight Race T epidemic of corn (*Zea mays* L.) in the United States in 1970 all were due to insufficient biodiversity for important genes in the affected crops (Large, 1962; Matossian, 1989; National Research Council, 1972). Each crop was uniformly susceptible to a critical disease—it was genetically vulnerable. As a consequence, yield and crop quality were drastically reduced. (All of these diseases were fungi. Wheat rust is caused by several species of *Puccinia,* ergot by *Claviceps purpurea* (Fr.: Fr.) Tul., potato late blight by *Phytoph-*

thora infestans (Mont.) de Bary, and Southern Corn Leaf Blight Race T by *Cochliobolus heterostrophus* (Drechs.) Drechs., Race T.)

Box 3

Each year the early Romans sacrificed a red dog to the corn god Robigus in hopes that he would accept it with pleasure and therefore protect their wheat crop from the destructive red rust that usually appeared at about the time of the ceremony, on April 25. Their wheat crop was genetically vulnerable (Large, 1962).

II. PLANT BREEDING, CROP VARIETIES, AND BIODIVERSITY

A. Applications of Biodiversity in Crop Agriculture

1. Genetic Diversity in Place—Spatial Diversity

Box 4

Genetic diversity in place (in the field) decreases the odds of catastrophic crop failure and starvation for subsistence farmers.

Early farmers found that one way to protect themselves from the problems of catastrophic disease, insect, and environmental problems was to maintain some degree of biodiversity among and within their crop species. Planting several kinds of crops gave assurance that at least some of them could survive any specific attack of disease, insects, or bad weather. Planting several varieties of a given crop increased assurance that some varieties of that crop could be harvested. Furthermore, genetic variability (seen by the farmers as phenotypic variability, i.e., visual variability) within a crop variety gave some assurance that at least some plants in the variety would be resistant to a specifically targeted, virulent strain of disease or insect. Each of these kinds of biodiversity (genetic diversity) decreased the odds of starvation. They provided "genetic diversity in place" (spatial diversity), that is, biodiversity in the field and in the countryside during the growing season.

2. Genetic Diversity in Time—Temporal Diversity

Box 5

Commercial farmers in industrialized countries such as the United States depend more heavily on genetic diversity in time than on genetic diversity in place.

Farmers have always, sometimes without even trying, relied on a second kind of biodiversity, "genetic diversity in time" (temporal diversity). Seed planted in successive seasons was never quite the same as that from the previous year, for the combined onslaughts of disease, insects, and weather eliminated many susceptible plants. Only the survivors could produce seed, and their genotypes reflected tolerance to the biological and weather-related constraints of the previous growing season. Varieties changed continually, even in the absence of purposeful selection by the farmers. The sequential genetic changes represented continuing adaptation to prevailing constraints to yield.

Farmers made conscious selection for changes in their varieties as well, sometimes for desired grain or plant traits, sometimes for yield per se. They also looked at their neighbors' varieties in the next village, the next valley, or even farther afield if transportation and communications allowed. Varieties were traded, experimental plantings were made, and new varieties often replaced older traditional varieties if farmers judged that the new ones did a better job of meeting their needs. This practice, of purposely producing "genetic diversity in time" via selection or variety replacement, is still common among farmers in areas that are little affected by professional plant breeding.

Today's professional plant breeders and modernized farmers work as a team to employ genetic diversity both in place and in time to provide protection against pest and weather problems, but they use tactics that are different from those employed by yesterday's farmer/breeders. They take advantage of modern technology and newer kinds of societal organization (Duvick, 1984). Planted areas are larger, crops are more uniform, replacements are more frequent, and sources of new genetic diversity come not from the neighboring village but from anywhere in the world. Plant breeders ceaselessly produce new varieties with increased yield, needed new kinds of pest resistance, adaptation to new cultural regimes, or better tolerance to bad weather (Duvick, 1996). Farmers continually experiment with the new varieties, choose the best of them, and plant them extensively, often so extensively that the stage is set for another explosive round of increased pest problems or too-uniform susceptibility to unexpected weather problems. Following such upsets, farmers immediately switch to other varieties, choosing from the ever-present catalogs of new introductions. Modern communication and transportation facilitate such rapid changes. They would not have been possible in earlier times.

3. Genetic Diversity in Reserve

Box 6

Genetic diversity in reserve has several layers, each layer farther away in time from use on the farm, but also broader in the amount of biodiversity that it can provide.

There is a caveat—a warning—about reliance on genetic diversity in time. It will work only when there is sufficient "genetic diversity in reserve" to supply a continuing stream of new varieties. The farmer-breeders of earlier times used their heterogeneous varieties as a reserve of genetic diversity as well as for active food production. They had no other ready reserve, no seed company offering a diverse supply of new varieties each year. But these subsistence farmers could also look farther afield to varieties grown by neighbors, and sometimes they might select from progeny of an occasional outcross to wild or weedy crop relatives. So, ready reserves and backup reserves of biodiversity were available and were used by subsistence farmers, but the scope usually was limited to the immediate locale and the current season's crops. Many of today's subsistence farmers still use these methods of conserving and using genetic reserves; they have no other options.

Today's commercial farmers, served by full-time professional plant breeders, have much broader reserves. The breeders have a supply of hundreds or even thousands of experimental varieties for each crop and each adaptation zone. They are grown in performance trials every year, enabling breeders to supply well-tested new varieties when needed. These are "front-line reserves." The plant breeders have other reserves, farther to the rear. These are "breeding pools" of materials in the process of hybridization and selection for production of new experimental varieties.

Behind the breeding pools reserve is a very broad, highly diverse set of varieties and wild and weedy relatives from all around the world, stored in long-term seed banks such as the regional USDA plant introduction stations and the National Seed Storage Laboratory at Ft. Collins, Colorado. Table I lists some of the world's largest seed banks and shows the size of their collections.

Finally, available but not easily accessed are the highly diverse materials now growing on farms and in the wild in all parts of the world. Collecting expeditions can gather materials from these sources, paying especial attention to "centers of diversity" for a given species, and make them available to plant breeders for long-term breeding programs with the ultimate goal of adding to the biodiversity of materials offered to farmers as new crop varieties. Table II lists major and minor centers of diversity worldwide and notes some of the important crops in those centers.

Genetic diversity in reserve thus has several layers, each layer farther away in time from use on the farm, but also broader in the amount of biodiversity that it can provide. As new varieties are introduced, become popular, and then eventually (and always) show unsus-

TABLE I

Estimates of Holdings in Major Plant Germplasm Banks[a]

Country/center	Crop	Number of accessions
Countries		
USA	All	557,000
China	All	400,000
Russia	All	325,000
India	All	76,800
Japan	All	60,000
Centers		
IRRI	Rice	86,000
ICRISAT	Sorghum, millet, chickpea, peanut, pigeon pea	86,000
ICARDA	Cereals, legumes, forages	77,000
CIMMYT	Wheat, maize	75,000
CIAT	Common bean, cassava, forages	66,000
IITA	Cowpea, rice, root crops	40,000
AVRDC	Vegetable crops	38,500
CIP	Potato, sweet potato	12,000

[a] Adapted from Board on Agriculture and National Research Council (1993). *Managing Global Genetic Resources: Agricultural Crop Issues and Policies.* National Academy Press, Washington, D.C. "Accessions" typically are individual crop varieties. IRRI, International Rice Research Institute; ICRISAT, International Crops Research Institute for the Semi-Arid Tropics; ICARDA, International Center for Agricultural Research in the Dry Areas; CIMMYT, Centro Internacional de Mejoramiento de Maíz y Trigo; CIAT, Centro Internacional de Agricultura Tropical; IITA, International Institute for Tropical Agriculture; AVRDC, Asian Vegetable Research and Development Center, C.I.P., Centre Internacional de la Papa.

TABLE II

Centers of Diversity for Some of the World's Major Food Plants[a]

Center of diversity	Important food plants native to the center
Major Centers	
1. Ethiopia	Barley, castor bean, coffee, flax, onion, sorghum, wheat
2. Mediterranean	Asparagus, beets, cabbage, lettuce, grapes, oats, olive
3. Asia Minor	Alfalfa, barley, carrots, cherry, date palm, pear, wheat
4. Central Asiatic	Apple, apricot, cantaloupe, cotton, grapes, onion, peas
5. Indo-Burma	Cotton, cucumber, eggplant, lemon, orange, pepper (black), rice
6. Thailand, Malaysia, Indonesia	Banana, breadfruit, coconut, ginger, grapefruit, sugarcane, yam
7. China	Buckwheat, peach, radish, rhubarb, rice, soybean, tea
8. Mexico–Guatemala	Bean (common), corn, cacao, cotton, pepper (red), squash, sweet potato
9. Peru–Ecuador–Bolivia	Bean (lima), cacao, cotton, guava, papaya, potato, tomato
Minor Centers	
10. Southern Chile	Potato, strawberry (Chilean)
11. Brazil–Paraguay	Brazil nut, cashew, cassava, peanut, pineapple
12. United States	Sunflower, blueberry, cranberry, Jerusalem artichoke

[a] Adapted from G. Wilkes (1983). Current status of crop plant germplasm. *CRC Critical Reviews in Plant Science* 1, 133–181. These centers are the place of origin and/or early domestication of most of the world's food crops. They still contain a very large share of the global genetic diversity for those crops.

pected weakness, genetic diversity in time supported by genetic diversity in reserve is used to protect and aid farmers in their constant struggle to increase yields in sustainable fashion.

B. Precarious Balance: Monoculture and Polyculture

Box 7

> Even the most diverse polycultures provide only a fraction of the biodiversity present in "natural" ecosystems, but they nevertheless can make the difference between starvation and plenty in some farming systems.

Crop agriculture, from its beginning and continuing to the present time, entails a continual struggle to balance the imperative of reducing biodiversity to maximize yield with the imperative of increasing biodiversity to maximize stability. Farmers must constantly balance their commitment to grow only the most productive varieties of a few favored crops with their need to maintain sufficient biodiversity to protect against unexpected failures of the "best" varieties and the most favored crops.

The terms "monoculture" and "polyculture" sometimes are used to describe the two extremes of crop biodiversity. Monoculture refers to growing a single, usually uniform, variety over large expanses. Polyculture refers to the practice of growing several crops in rotation or in mixture, often in smaller areas and often with some variability within or among varieties. Although both terms are used with remarkable lack of precision and consistency, they do express the fact that crop culture and crop genotypes can vary widely across the global farming community.

Usually overlooked, but nevertheless unavoidably true, is the fact that even the most diverse polycultures provide only a fraction of the biodiversity present in "natural" ecosystems—ecosystems that are relatively untouched by human intervention. A polyculture cropping system, with two or three crops in one field and a certain amount of variability of genotype in each crop, is much closer to a monoculture than to any relatively pristine ecosystem, excepting perhaps a few deserts, tundras, or salt flats. Even relatively simple natural ecosystems, such as the tallgrass prairies of the American Midwest, contain hundreds of highly variable plant species compared to the half dozen or more that might be found in one of the more diverse agricultural polycultures.

But comparisons aside, the small amount of biodiversity provided to farming systems by biodiversity in place (spatial) and biodiversity in time (temporal) is essential and beneficial (Tilman *et al.,* 1999). When properly managed it can make the difference between famine and plenty, or between economic ruin and a satisfactory income. The particular kinds and amounts of needed biodiversity vary by crop, by farming system, and by the norms and needs of societal and economic systems. This article reviews some of these issues, concentrating on the use of biodiversity for the development of field crops grown by commercial farmers in the United States. Its content draws heavily on an informal survey of a score of experienced professional plant breeders.

C. Contributions of Biodiversity to Today's Crop Varieties

1. Increased Yield

Box 8

As the relationship of base-broadening germplasm to that which is being improved decreases, so do the odds of producing successful new varieties, yet the odds increase that the infrequent successes will have a significant advantage in yield or other important traits.

Biodiversity in the form of genetically diverse breeding materials is essential for increasing the yielding ability of farm crops. Yield is the key trait for breeders of commodity crops such as wheat, soybeans (*Glycine max* (L.) Merrill), and corn. New varieties must yield more than those they are to replace, and they must do so repeatedly, season after season. Development of new varieties from the progeny of crosses of high-yielding varieties gives dependable but small increases in yield. The high-yield varieties often are related to each other. But through trial and error, breeders have found that the infusion of new germplasm from unrelated families can give large, stepwise increases in yielding ability. Sometimes the improvement in yield clearly is due to the purposive addition of traits that were lacking in older varieties, such as tolerance to heat and drought or resistance to a prevalent disease. But often the improvement is unpredicted and its basis is unknown. The only constant from the breeders' point of view is that one continually must broaden the germplasm base of any breeding population in order to keep the curve of yielding ability on an upward slope (Duvick, 1984).

Both intuition and scientific knowledge are used in choosing materials for the broadening of the genetic base. It is fairly easy to identify base-broadening materials that will have very little utility for variety improvement, but it is not easy to identify materials with high probability of giving large increases in yielding ability. Some breeders outcross only to elite but unrelated germplasm that is already adapted to the local growing conditions. Others also work with elite germplasm of widely different adaptation, and a few breeders infuse germplasm from seemingly undesirable sources, such as outdated farmer varieties or wild relatives of the crop species.

All of these methods work, but with widely varying rates of success. A rule of thumb is that as the relationship of the base-broadening germplasm decreases so do the odds of producing successful new varieties, but at the same time the odds increase that the infrequent successes will give large advantages in yield.

2. Hybrid Vigor

Box 9

Germplasm from diverse sources can give new kinds of heterosis and thus hybrids with novel improvements, including greater yield and stability of yield.

Biodiversity is essential for maintaining and increasing hybrid vigor. Field crops that are bred and sold as hybrids—such as corn, sorghum (*Sorghum bicolor* (L.) Moench), and sunflower (*Helianthus annuus* L.)—depend on the phenomenon of heterosis, or hybrid vigor, for part of their yielding ability. When heterosis is high, crosses between two inbred lines yield substantially more than either parent, even though the inbreds themselves may have been selected for high yield per se. (Inbred lines are uniform products of several generations of self-pollination.) Breeders sort inbred lines into "heterotic groups" based on performance of the inbreds in hybrid combination. Hybrids made by crossing inbreds from contrasting heterotic groups have more hybrid vigor, and in general higher yield, than hybrids made by crossing inbreds within a heterotic group. Breeders search continually for new heterotic groups, hoping to find new cross-group combinations with increased yield. Germplasm from diverse sources also can give new heterotic groups with useful traits that were not found in the earlier heterotic groups.

3. Disease and Insect Resistance

Box 10

Biodiversity both helps and hinders plant breeders as they develop new varieties that have improved resistance to pests and pathogens.

To plant breeders, nothing is more self-evident than their dependence on biodiversity to stay ahead of (or catch up with) continually changing and challenging problems with disease and insect pests. All too often, new biotypes of an insect pest or new races of a destructive fungus will overcome the genetic resistance in currently-grown varieties and all of their close relatives. Breeders then need to look for new resistance genes in far-away sources, particularly those with high genetic diversity. The broad range of genotypes available in traditional farmer varieties is an especially rich source of resistance genes.

a. Specific Resistance

For many crops such as soybeans and sorghum, numerous individual genes from exotic germplasm in all parts of the world will give high levels of resistance to diseases such as *Phytophthora* root and stem rot (from *Phytophthora sojae*; M. J. Kaufman & J. W. Gerdemann) or to insects such as greenbug (*Schizaphis graminum* Rondani). Such kinds of highly targeted resistance are strikingly effective, but they usually do not last for many seasons, for the pest species typically has a reservoir of genetic variants that can overcome the single genes for resistance. The variants multiply and soon the "resistant" crop varieties are susceptible. Such specific resistance, usually given by one or a few genes of large effect but often with short effective life-time, is sometimes called "vertical resistance" (van der Plank, 1963).

b. Durable Resistance

Heritable pest resistance also can be imparted by complex assemblages of genes, and insects and disease organisms are rarely able to overcome such kinds of resistance. The resistance is longer-lasting, and often is termed "durable" by plant breeders and others concerned with the ravages of disease and insect pests on crop plants. It also has been called "horizontal" resistance in contrast to single-gene "vertical" resistance (Simmonds, 1991; van der Plank, 1963).

A disadvantage of durable resistance provided by complex assemblages of many genes is that breeders cannot easily move the assemblages from one plant or one variety to another. The gene assemblages tend to disassociate during the successive segregating generations that are needed to develop a variety, or they tend to bring in associated ("linked") traits from the donor parent that detract from the utility (often the yield) of the elite material that breeders wish to improve. But advantages of durable resistance are so great that many breeders have devoted much time and energy to incorporating the best gene combinations into their breeding stocks (Simmonds, 1993). Very often the best sources of multi-gene durable resistance are unadapted exotics such as farmer varieties and wild or weedy relatives of crop species. In these cases problems with linkage to undesirable traits are accentuated. Thus biodiversity both helps and hinders the efforts of breeders to impart durable pest and pathogen resistance to new varieties of crop plants.

4. Broad Adaptation

Box 11

Use of diverse germplasm sources as parental breeding materials increases the odds that a new variety will have multiple kinds of effective genetic defense systems. Such "internal biodiversity" is the foundation for broad adaptation.

Theorists disagree about the utility of broad adaptation and its relationship to biodiversity. If a single variety fares well over a broad geographical range such as the full expanse of the U.S. Cornbelt, it offers more opportunity for the large-scale multiplication of disease or insect pests that can overcome its defenses. Large-scale plantings of a single variety increase the chances of genetic vulnerability.

On the other hand, it is self-evident that varieties could not have broad adaptation unless they contained multiple kinds of effective genetic defenses, able to counter the multiplicity of environmental and biological constraints of a wide range of localities. Broadly adapted varieties would appear to have greater internal biodiversity, that is, greater diversity of useful genetic systems in each plant.

Breeders say that varieties with maximum amounts of tolerance to the full range of environmental and biological stresses that could be expected in any specific locality are, therefore, broadly adapted as well. Selection for stability of performance over a variety of seasons in a single locality tends to produce varieties with "pre-adaptation" to a variety of localities, and such varieties often become popular in a much larger region than had been envisioned. It then is up to farmers and seed companies to decide how widely such varieties will be planted, and at what ratios in comparison to genetically different varieties with (perhaps) lower yields or lower stability of performance. Once again, the urge to maximize yields conflicts with the need to optimize biodiversity.

Breeders also agree that diverse germplasm sources are most likely (after a long and often frustrating period of hybridization and selection) to give varieties with maximum amounts of internal biodiversity, that is, with multiple defense systems that are the basis for broad adaptation. Biodiversity at the varietal level is thereby internalized to biodiversity at the genomic level.

FIGURE 1 A diverse assemblage of fruits and flowers of wild and domesticated tomatoes. Such diversity was the basis for the development of tomato as a multi-use crop. It also will be the foundation for improvements in the crop in years to come. (Photograph courtesy of North Central Regional Plant Introduction Station, Ames, Iowa.) See also color insert, this volume.

5. New Traits

Box 12

Unlikely sources of diverse germplasm can provide valuable new traits.

In horticultural crops, and increasingly now in field crops, new traits may have great economic value. Almost invariably, breeders must go to genetically diverse sources for such new traits. Wild species related to tomato (*Lycopersicon esculentum* Mill.) contributed genes that enhanced the soluble solids and sugar content of tomatoes, traits that are of large commercial importance for certain uses of tomato. (Figure 1 illustrates the impressive amount of phenotypic diversity that exists among wild and domesticated tomatoes.) At the present time, an alliance of public and private corn-breeding organizations is introgressing tropical maize (corn) into elite temperate germplasm with the intention of finding new levels of oil content, or other new grain quality traits with commercial utility. Similarly, glandular trichomes (specialized leaf hairs) from a related wild species promise to give a new kind of insect resistance to potatoes.

6. Pleasant Surprises

Box 13

Linked genes often cause problems but sometimes they provide pleasant surprises.

One of the pleasures of plant breeding is to breed for one trait and discover, fortuitously, that the act of breeding has produced a new entirely unexpected but highly desirable trait. This was the case with cucumber (*Cucumis sativus* L.) when a breeder outcrossed to a variety from China in order to incorporate its virus resistance into adapted germplasm, and discovered that the variety also contributed an entirely new kind of color pattern to the fruit. The new pattern, a uniform fruit color, proved to be highly desirable for commercial use when enhanced with dark green background color. It now is incorporated into virtually all of the commercial cucumber varieties.

In another example, a corn breeder in Tennessee used germplasm from Cuban open-pollinated varieties to improve ear worm (*Heliothis zea* Boddie) resistance in hybrids for use in the Midsouth. An inbred line from that breeding program fortuitously had excellent resistance to a new strain of virus that swept across the southeastern United States. The resistance probably was derived from the Cuban germplasm. The inbred became the basis for a series of successful virus-resistant hybrids that were adapted to the Southeast. These examples demonstrate that gene linkage, which is often a problem, sometimes can provide pleasant surprises for plant breeders and the farmers they serve.

7. Transgenes and Key Biological Functions

Box 14

In a sense, biotechnology has closed the circle of biodiversity as described by E. O. Wilson. The search for new kinds of useful biodiversity now can encompass "the totality of hereditary variation in life forms, across all levels of biological organization."

The advent of biotechnology in the aid of plant breeding led some to say that plant breeders would no longer depend on diversity within the crop species for new advances. It was believed that genetic transformation (transfer of genes without benefit of sexual hybridization) could bring in powerful new genes as needed from any one of the biological kingdoms. Naturally occurring genetic diversity as found within a crop species would no longer be needed, or at least would become secondary in importance and use.

However, after 20 years of experience it now appears that biotechnology may have increased rather than decreased the value of biodiversity within a crop species and its close relatives. Although some genes of great promise have been moved from far afield into crop plant varieties (e.g., inserting insect resistance genes from bacteria to maize) and more will soon follow, the total number is not large.

On the other hand, genetic transformation has enabled biologists to learn more than was ever thought possible about how genes are regulated. Molecular biologists are learning how to increase or decrease the rate of formation of gene products (such as enzymes), and to control the time and place at which a gene is turned on. They increasingly are able to identify most of the genes in a crop species and are beginning to identify their functions. The goal is to find genes and gene systems of key importance in controlling important traits such as tolerance to heat and drought, disease and insect resistance, or time of flowering, and then to enhance their operation.

A great aid to acquiring this new knowledge is the ability to search through the genomes of a genetically diverse assemblage of local crop varieties, then to look farther afield to varieties adapted to other parts of the globe, then to wild or weedy relatives of the crop species, and eventually to unrelated species and even organisms from other biological kingdoms. Comparisons of like genes or like genetic systems at each level add incrementally to understandings of how the assemblages of genes function in a particular crop and give hints about how to make them function better. Biotechnology gives plant breeders the ability to bring in foreign genes from any part of the world of nature, but more importantly it helps them to improve the action of native genes in their original settings, based on insights gained from the study of similar genes in other organisms. In a sense then, biotechnology has closed the circle of biodiversity as described by E. O. Wilson by allowing plant breeders to manipulate "the totality of hereditary variation in life forms, across all levels of biological organization." Biodiversity can contribute to agriculture more strongly than ever, thanks to biotechnology.

8. New Cultural Systems

Box 15

In addition to inventing crop varieties, the first farmers invented ways to grow them in combination with each other. We have inherited and still use the results of their experiments, often adapted in ways that disguise their origins.

At the other end of the spectrum from molecular biology is the biology of assemblages of whole plants and animals, as well as microbes, and how they interact with their environment. Although some people believe that the term "ecology" first described this holistic concept, the fact is that the term "farming" in whatever language has a precedence of about 10,000 years as a descriptor of at least one class of ecological systems. Farming is all about the biology of assemblages of plants, animals (including humans), and microbes and how they interact with their environment.

The first farmers also tested and developed ways to grow different crop varieties in combination with each other. Although their primary aim was to simplify farming and grow only the crops desired, many kinds of cropping systems were designed, each intended to give satisfactory food production for a particular locale and life-style. We have inherited the results of those experiments, the failures and successes of these first farmers. Some of the methods still work for present needs, some have been reduced in importance or even largely abandoned, and others have been adapted in ways that disguise their origins.

a. Rotations

Crop rotations were essential before the advent of chemical-intensive agriculture. Farmers used rotations to "rest the soil," to help in weed control, and, perhaps without clear understanding, to help control disease and insect problems. Biodiversity among crop species had obvious utility to the early farmers. Crop varieties were selected to fit rotations, and rotations were designed to fit the crop varieties.

Crop agriculture in the United States still uses rotations, but often they are greatly simplified, as in the common corn/soybean two-year rotation in the American Midwest. The purpose of rotations is still the same, namely, to enable successive crops to benefit from those that preceded them and, in turn, to benefit those that follow. Soybeans help corn to escape damage from the corn rootworm beetle (*Diabrotica* spp.) by providing a year in which the field is not infested with rootworm eggs and subsequent root-chewing larvae. In this way, farmers can avoid the use of chemical pesticides to control corn rootworm.

But biodiversity among populations of rootworm beetles has enabled some of them to wait an extra season before hatching. They emerge just in time for the corn crop following the soybean rotation. Other populations have developed the ability to feed and lay eggs in soybean fields, and they, too, produce larvae just in time for the next year's corn crop. Farmers, plant breeders, and entomologists now are working to devise new rotational schemes or new kinds of plant varieties that might put the rootworm beetles at a new disadvantage. Biodiversity once again both helps and hinders agriculture, depending in this case on whether it is in the crop or in the pest. To use biodiversity to profit humans (and not rootworms) requires keen biological insights and knowledge of ecological interactions at many levels. Use of rotations to provide beneficial biodiversity is no simple matter.

There can be no mistake, however, that the utilization of specifically designed rotations adapted to today's technology and economy can be cost saving and chemical saving, while also maintaining or increasing the yields and quality of the product. The principle works worldwide. In India's Punjab State, for example, rice (*Oryza sativa* L.) yields following potato and maize were 36% higher than when rice followed wheat, a typical rotation in the Punjab.

b. Polycultures

Subsistence farmers, especially in the tropics, use polycultures routinely; that is, they grow mixtures of crops planted together in various patterns. The mixed crop plants complement each other by varying the time at which they draw on water or nutrients from the soil, or by adding beneficial nutrients for use by other crops [e.g., nitrogen-fixing beans (*Phaseolus vulgaris* L.) planted with corn], or by providing needed shade to other crops. Farmers in medieval Europe often grew mixtures of wheat and rye ("méteil") as insurance against the failure of wheat, which was the more delicate but more preferred of the two crops (Neveux *et al.*, 1975).

Some researchers are experimenting with polycultures for use in modern commercial farming. They intend to produce special types of grain crops that can be grown in mixtures. Such polycultures could reduce the need for chemical fertilizer, increase the stability of production, provide non-chemical protection against disease and insect pests, and reduce soil erosion if the polycultures are composed of perennial plants. One group of researchers uses the tallgrass prairie ecosystem as a model. They intend to copy the salient features that give it stability, productivity, and, not least, the capacity for reducing soil erosion (Soule and Piper, 1992). The polyculture breeders intend to develop crop mixtures with useful degrees of intra- and inter-species diversity that also produce crop products that will meet standards of yield and quality for the needs of today's farmers and the consuming public.

Researchers agree that development of polyculture crops presents large challenges as compared to making improvements to one crop at a time. To simultaneously select several crops for individual and mutual enhancement when grown in one polyculture requires analyses that will test skills in statistics as well as in agronomy and plant physiology. But simply to attempt such a project will give benefits in knowledge that could well exceed the value of the final product.

For example, plant-rows (each plant-row is a monoculture of seed from a single plant) of wild species have shown great variability in resistance to certain fungal diseases. Some plant-rows show no symptoms whereas others are highly susceptible. This behavior contrasts with the fact that individual plants of the species show no or little damage in normal prairie settings. It would seem that protection afforded by biodiversity at the species level (both within and among species) was lost when individual genotypes were planted as single-row monocultures. An alternative explanation is that diseased plants are hard to spot in natural prairie settings, and plant-rows simply make it easier to categorize the variation that exists in the wild.

The plant-row problems with disease testify to the great advances that had to be made by the first farmers when they put out similar isolated plantings of a single species in the first stages of domestication. The crop was deprived of intra- and inter-species protection and the farmers had to select (even if without special intent) for genetic systems that were powerful enough to give resistance without this protection.

c. Multilines

Modern science has shown that carefully planned mixtures or blends of different selections of the same variety, differing only in major genes for disease resistance, can provide reasonably durable protection against diseases. In oats (*Avena sativa* L.), for example, a series of resistance genes is backcrossed into a single variety, producing numerous strains, each with a different resistance gene. Each gene carries resistance to a specific race of a fungal disease such as crown rust (*Puccinia coronata* Cord), and each gene carries a different kind of resistance. The several strains are then mixed and sown together as one crop, apparently homogeneous but in actuality genetically heterogeneous for resistance to the disease. Such a mixture is called a "multiline." As noted, its diversity of resistance genes will cover all known races of the disease, and essentially provides "genetic diversity in place." But as new races and new genes conferring resistance to them are found, new backcross strains of the oat variety can be produced and added to the original, thus breeders also can take advantage of "genetic diversity in time."

III. CONCLUSIONS

Plant breeding—the development of plant varieties—simultaneously exploits and enhances biological diversity.

Plant breeding exploits biodiversity. Modern plant breeding could not exist without recourse to a continuing supply of biologically diverse populations at the variety, species, and family levels, and now (thanks to biotechnology) at any level in the world of nature.

Plant breeding enhances biodiversity. Plant breeders of all kinds—full-time professionals, farmer-breeders, and dedicated amateurs—add to the stock of genetically diverse organisms by the continual production of new and genetically diverse plant varieties with new, ecologically diverse adaptations. Such increased biodiversity is minuscule when compared to that existing in relatively pristine ecosystems, but it is essential for productive and stable agricultural production.

Yet plant breeding has also caused problems when breeders and farmers ignore or misunderstand the ways in which biodiversity contributes to ecological balance and crop productivity. Excessive dependence on simple solutions such as single-gene disease resistance has led to problems.

Plant breeding has given the greatest benefit when its products and their users took advantage of the beneficial interactions that accrue among diverse organisms at each level of complexity from individual genes to the landscape. Much remains to be learned about ways to make the greatest use of biodiversity at each level (e.g., spatial, temporal, or reserve), about how one kind can substitute for another, and when it is best not to substitute. Plant breeding epitomizes the duality of humanity's interaction with biodiversity. We wish to alter and enhance it for the benefit of human needs and wants, but we also must avoid altering or depleting it in ways that destroy its benefits to us and to the world of nature—our home.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • AGRICULTURE, SUSTAINABLE • AGRICULTURE, TRADITIONAL • BREEDING OF ANIMALS • GENETIC DIVERSITY

Bibliography

Board on Agriculture and National Research Council. (1993). *Managing Global Genetic Resources: Agricultural Crop Issues and Policies.* National Academy Press, Washington, D.C.

Duvick, D. N. (1984). Genetic diversity in major farm crops on the farm and in reserve. *Econ. Botany* **38**, 161–178.

Duvick, D. N. (1996). Plant breeding, an evolutionary concept. *Crop Sci.* **36**, 539–548.

Large, E. C. (1962). *The Advance of the Fungi.* Dover Publications, New York.

Matossian, M. K. (1989). *Poisons of the Past: Molds, Epidemics, and History.* Yale University Press, New Haven, Connecticut/London.

National Research Council. (1972). *Genetic Vulnerability of Major Crops.* National Research Council, National Academy of Sciences, Washington, D.C.

Neveux, H., J. Jacquart, and E. L. R. Ladurie. (1975). *Histoire de la France Rurale, L'Age Classique.* 2. Seuil, Tours, France.

Simmonds, N. W. (1991). Genetics of horizontal resistance to diseases of crops. *Biol. Rev.* **66**, 189–241.

Simmonds, N. W. (1993). Introgression and incorporation strategies for the use of crop genetic resources. *Biol. Rev.* **68**, 539–562.

Smith, B. D. (1995). *The Emergence of Agriculture.* Scientific American Library, distributed by W. H. Freeman and Company, New York.

Soule, J. D., and J. K. Piper. (1992). *Farming in Nature's Image.* Island Press, Washington, D.C./Covelo, California.

Tilman, G. D., D. N. Duvick, S. B. Brush, R. J. Cook, G. C. Daily, G. M. Heal, S. Naeem, and D. Notter. (1999). *Benefits of Biodiversity.* Task Force Report, No. 133, February 1999. Council for Agricultural Science and Technology, Ames, Iowa.

van der Plank, J. E. (1963). *Plant Diseases: Epidemics and Control.* Academic Press, New York.

Wilson, E. O. (1994). *Naturalist.* Island Press/Shearwater Books, Washington, D.C./Covelo, California.

BUTTERFLIES

Philip J. DeVries
Center for Biodiversity Studies, Milwaukee Public Museum

GLOSSARY

aposematic Describing an organism that is rendered less susceptible to predation by advertising its obvious unpalatability.

Batesian mimicry A form of mimicry in which the target organism is rendered less susceptible to predation by its resemblance in morphology or coloration to a different species that is unpalatable.

cryptic Describing an organism that is concealed or obscured by the similarity of its appearance to the surrounding environment.

Müllerian mimicry A form of mimicry in which two or more unpalatable species resemble each other, with the effect that predators are more likely to avoid any species with this appearance.

myrmecophily Ability to form symbiotic associations with ants.

vibratory papillae Mobile, grooved, rod-like appendages arising from the distal edge of the first thoracic segment, used for communicating.

BUTTERFLIES LIKELY REPRESENT the most familiar and best known group of all insects. Within the context of human society, butterflies serve as centerpieces in educational media, they are used extensively in the arts, including nature and commercial advertising, and they are used as symbols for religious and social groups. Within the field of biology, studies on butterflies have been fundamental to the development of biogeography, behavior, coevolution, conservation, development, ecological genetics, evolution, global warming, mimicry, population ecology, sexual selection, speciation, symbiotic associations, and systematics. In summary, butterflies have been important to how we perceive biodiversity.

I. OVERVIEW OF BUTTERFLY TAXONOMIC DIVERSITY

Although butterflies may be the best known group of insects, our understanding of their taxonomic diversity has two fundamental weaknesses. The first regards the recent decline of professional butterfly taxonomists and species level revisions in the past 50 years. This requires most species diversity estimates (from family to genus) to be derived from a literature that is out-of-date. The second is that the number of families and subfamilies of butterflies varies among classifications because, despite current interest in phylogenetic systematic methods and analyses, the relationships within the major groups are

Encyclopedia of Biodiversity, Volume 1

unresolved, particularly within Nymphalidae and Lycaenidae. Such is the state of butterfly taxonomy. Nevertheless, the "true" butterflies (superfamily Papilionoidea) may be placed conservatively into four families (Papilionidae, Pieridae, Nymphalidae, and Lycaenidae) that together include between 12,900 and 15,819 species. One useful framework for organizing butterfly taxonomic diversity is P. R. Ackery's (1984) synthesis of butterfly classification, which forms the basis of the following synopsis:

Family Papilionidae (Swallowtails, Fig. 1): a group of 500–600 species in three subfamilies, distributed worldwide but with most species being tropical; adults medium to large sized, both sexes have six walking legs bearing nonbifid tarsal claws; most are brightly colored, may be Batesian or Müllerian mimics, all feed on flower nectar, and males drink at wet soil (Fig. 1); caterpillars are herbivores, body smooth without hardened spines, and possess extrusible glands (osmeteria) that are unique among butterflies; pupa typically with a silk girdle at third thoracic segment.

Subfamily Baroniinae: one species (*Baronia brevicornis*) endemic to Mexico; caterpillars unique among the family by feeding on Fabaceae.

Subfamily Parnassinae: 40–50 species, mainly in north temperate mountains; host plant families include Crassulaceae, Fumariaceae, and Zygophyllaceae.

Subfamily Papilioninae (Fig. 1): more than 500 species placed into three or more tribes; distributed worldwide; most species are tropical. Many species with long hindwing tails; many with sexes strongly dimorphic; many involved in mimicry; species range from palatable mimics (*Papilio* and *Eurytides*) to unpal-

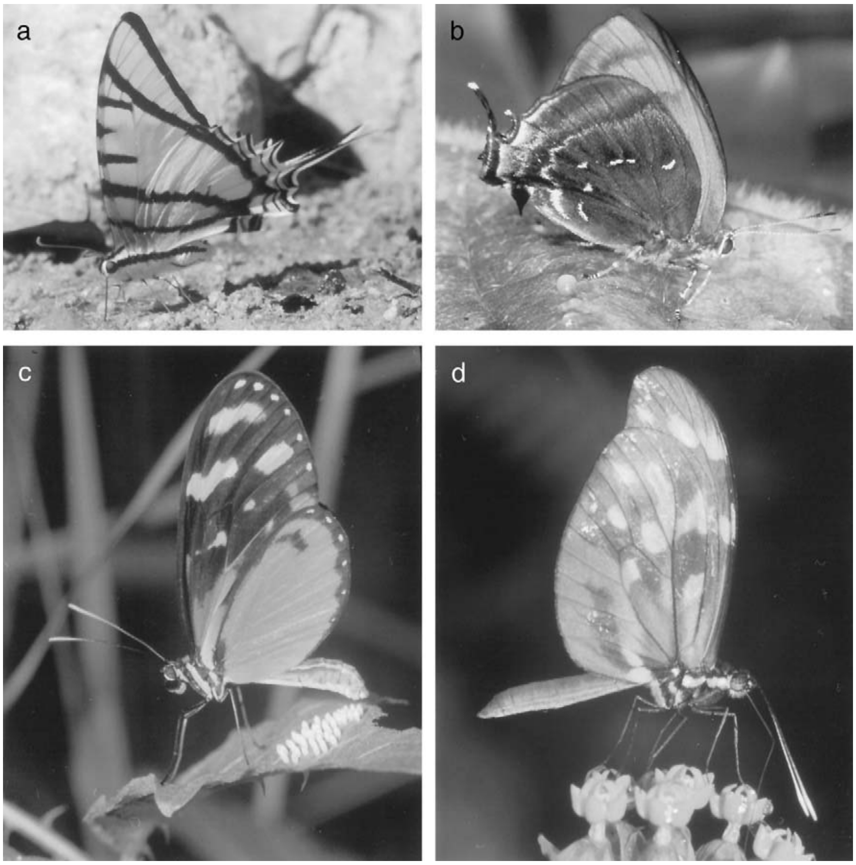


FIGURE 1 Adult butterflies. Clockwise from upper left: (a) newly eclosed male *Eurytides protesilaus* (Papilionidae) drinking from wet soil; (b) male *Theritas* nr. *hemon* (Lycaenidae: Theclinae) perching on a leaf; (c) male *Dismorphia amphiona* (Pieridae: Dismorphinae) visiting flowers; (d) female *Mechanitis isthmia* (Nymphalidae: Ithomiinae) ovipositing a cluster of eggs. [All photos copyright P. J. DeVries.]

atable models (*Parides, Pachliopta*, and *Troides*), and some are among the largest butterflies (*Trogonoptera* and *Ornithoptera*); host plant families include Annonaceae, Apiaceae, Aristolochiaceae, Fumariaceae, Hernandicaeae, Lauraceae, Magnoliaceae, Rosaceae, and Rutaceae.

Family Pieridae (whites, sulphurs, jezebels, cabbage butterflies; Fig. 1): a group of more than 1100 species in four subfamilies; adults small to medium-sized, both sexes have six walking legs and distinctly bifid claws; most species are white or yellow or orange (or combinations thereof) derived from pterin pigments, some with red and black patterning; adults feed on flower nectar (Fig. 1), and males visit wet soil; caterpillars are herbivores, body smooth without hardened spines, often covered in granulations; pupa suspended at 45° angle from substrate with silk girdle at the first abdominal segment.

Subfamily Pseudopontinae: one species (*Pseudopontia paradoxa*) endemic to West Africa; early stage biology unknown.

Subfamily Dismorphiinae: approximately 100 species, nearly all Neotropical. Most *Dismorphia* (Fig. 1) are astonishingly precise mimics of certain aposematic Nymphalidae (*Heliconius, Mechanitis*, and *Oleria*), and represent the original example from which Batesian mimicry theory was derived; host plants are in the Fabaceae.

Subfamily Pierinae (whites, jezebel, and cabbage butterflies): approximately 700 mostly tropical species, especially in Africa and Asia; some species have spectacularly bright, contrasting colors of white, red, and black (e.g., *Delias, Mylothris, Pereute*, and allies) suggesting unpalatability and warning coloration; some groups with dimorphic sexes in which females resemble unpalatable nymphalids and papilionids; some species undergo spectacular mass migrations (e.g., *Ascia, Appias*, and *Belenois*), and a few species are crop pests (*Pieris*); host plant families include Capparidaceae, Brassicaceae, Santalaceae, and Loranthaceae.

Subfamily Coliadinae (sulphurs): 400–600 species well represented in temperate and subtropical regions; generally yellow or white with short, thickly scaled antennae; common in open areas, some migrate in large numbers and have been recorded out at sea far from land (*Colias, Phoebis, Catopsilia, Aphrissa*, and *Eurema*); many species with patterns visible only in the ultraviolet.

Family *Lycaenidae* (hairstreaks, blues, coppers, and metal marks; Fig. 1): a group of 6000–6500 mostly tropical species in 10 subfamilies (the total number may change with future study), and accounts for nearly 50% of all butterflies; most species are small to very small, both sexes have six walking legs (except Riodininae), and frequently with alternating black and white bands on the antennae; the group displays a tremendous diversity of form, color, and life histories; adults feed on flower nectar, fruits, carrion, and honeydew, and a few species do not feed as adults; some African and Southeast Asian groups are involved in mimicry complexes; caterpillars most often slug-like, without hard spines or projections, but the Riodininae shows an extensive variety of form; as a group, lycaenids have the widest diet breadth of all butterflies, and depending on the group caterpillars may feed on plants, other insects, or insect secretions; many caterpillars form intimate and complex symbiotic associations with ants and produce acoustical calls similar to ant calls, secretions that are harvested by ants, and chemicals that alter ant behaviors; pupae are typically round, seed-shaped, unadorned with projections, and may produce clicking or whirring sounds when stimulated.

Subfamily Riodininae (metalmarks): 1200–1400 species that are almost exclusively Neotropical; arguably the most diverse of all lycaenoid groups with respect to adult and larval forms, and they are often treated as a distinct family (Riodinidae) divided into five subfamilies; adults feed primarily on flower nectar, some drink at wet soil, others at carrion; overall their life histories are poorly known, but as a group riodinid caterpillars appear to be mainly herbivores (*Euselasia, Mesosemia, Ancyluris, Metacharis, Mesene, Symmachia, Emesis, Anteros*, and *Helicopis*), with a few carnivorous species on Homoptera (*Alesa* and *Setabis*); about one-third of the riodinids form symbioses with ants and have vibratory papillae as sound producing organs [*Thisbe, Audre, Lemonias, Synargis, Nymphidium*, and *Setabis* (Fig. 2)]; riodinid myrmecophily is entirely Neotropical; host plant families include Araceae, Asteraceae, Bromeliaceae, Bombacaceae, Cecropiaceae, Clusiaceae, Dilleniaceae, Euphorbiaceae, Fabaceae, Lecythidaceae, Loranthaceae, Malpighiaceae, Marantaceae, Melastomataceae, Myrtaceae, Orchidaceae, Rubiaceae, Sapindaceae, and Zingiberaceae plus mosses, liverworts, and lichens.

Subfamily Styginae: one enigmatic species (*Styx infernalis*) that is endemic to the Peruvian high Andes; the early stage biology is unknown.

Subfamily Lipteninae: a small African group of 30–40 species; some are Batesian mimics of aposematic nymphalids (*Mimacraea* and *Ornipholidotos*); caterpillars feed on lichens and microscopic fungi (*Durbania* and *Deloneura*).

Subfamily Poritiinae: a small group restricted to the Oriental regions; caterpillars may be gregarious, feed

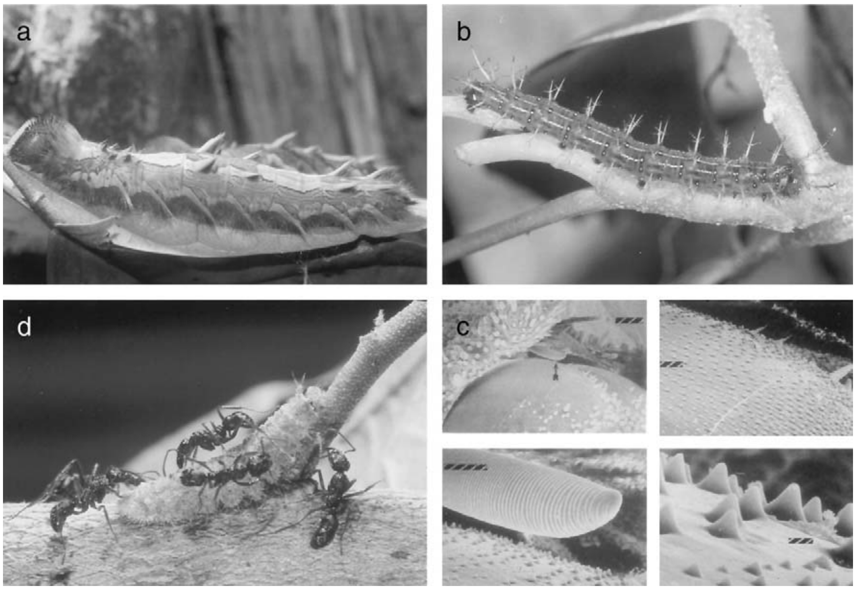


FIGURE 2 Butterfly caterpillars. Clockwise from upper left: (a) final instar caterpillar of *Morpho achilles* (Nymphalidae: Morphinae); (b) final instar caterpillar of *Tigridia acesta* (Nymphalidae: Nymphalinae); (c) Four frames showing details of the call production mechanism of *Thisbe irenea* (Lycaenidae: Riodininae). Upper left—top of caterpillar head, a vibratory papilla (arrow) and anterior margin of first thoracic segment. Upper right—granulations on top of caterpillar head. Lower left—detail of vibratory papilla. Lower right—head granulations at high magnification; (d) final instar caterpillar of *Thisbe irenea* (Lycaenidae: Riodininae) being tended by ants. [All photos copyright P. J. DeVries.]

on plants in the Fagaceae (*Poritia*), and may not associate with ants.

Subfamily Liphyrinae: a small group found in Africa, Australia, and Asia; adults have the proboscis partly or entirely atrophied; depending on the species, caterpillars feed entirely on insects (Homoptera) or on ant brood within ant nests (*Liphyra* and *Aslauga*).

Subfamily Miletinae: a moderate-sized group with most species in tropical Africa and the Orient and one in North America; caterpillars mainly feed on nymphs of Homoptera (*Feniseca, Spalgis*, and *Allotinus*), with a few species feeding on secretions of Homoptera (*Allotinus*) or ant regurgitations (*Thestor*); adults feed mainly on honeydew secretions of Homoptera.

Subfamily Curetinae: approximately 40 species restricted mainly to tropical Asia, all in the genus *Curetis;* caterpillars with tentacle organs as conspicuous, rigid cylindrical tubes; caterpillars feed on Fabaceae and have loose associations with ants.

Subfamily Theclinae (hairstreaks; Fig. 1): this very large and diverse group is found worldwide, but most species are tropical; caterpillars are mainly herbivores, some carnivores on Homoptera, and many associate with ants, but the life histories of most species remain unknown; host plant families include Anacardiaceae, Annonaceae, Asteraceae, Bromeliaceae, Clusiaceae, Cycadaceae, Euphorbiaceae, Fabaceae, Fagaceae, Geraniaceae, Lecythidaceae, Loranthaceae, Malpighiaceae, Malvaceae, Melastomataceae, Meliaceae, Myrtaceae, Oleaceae, Orchidaceae, Sapindaceae, Sapotaceae, Solanaceae, Sterculiaceae, Ulmaceae, and Verbenaceae.

Subfamily Lycaeninae (coppers): a group of 20–40 species found mainly in temperate regions, a few are tropical; caterpillars are herbivores on Polygonaceae, and some form loose associations with ants.

Subfamily Polyommatinae (blues): a large group with a worldwide distribution, most are pale, reflective blue above; caterpillars appear to always associate with ants, and their diets range from herbivores to carnivores, and some (*Maculinea*) feed on ant larvae and are clearly complex parasites and predators within ant nests; host plant families include Crassulaceae, Euphorbiaceae, Fabaceae, Lamiaceae, Myrsinaceae, Oxalidaceae, Primulaceae, Rhamnaceae, Rutaceae, Santalaceae, Sapindaceae, Saxifragaceae, Selaginaceae, Sterculiaceae, Verbenaceae, Zingiberaceae, and Zygophyllaceae.

Family Nymphalidae (brush-footed butterflies; Figs. 1 and 2): a group of 4800–6000 species in 14 subfamilies embracing a prodigious variety of forms and sizes, with both sexes having four walking legs; the forelegs are greatly reduced (hence "brush-footed"); adults may be dull brown (Satyrinae and Nymphalinae), brightly colored (Nymphalinae and Brassolinae), brilliantly iridescent due to physical properties of the wing scales (Morphinae and Apaturinae), or transparent (Ithomiinae and Satyrinae); some groups are entirely palatable, others highly distasteful, and some (Danainae, Ithomiinae, Heliconiinae, and Acraeinae) are extremely important mimetic models; some species (Charaxinae and Nymphalinae) mainly inhabit tropical forest canopies; adults may feed on flower nectar, pollen, rotting fruits, carrion, or do not feed at all; the caterpillars may bear many spines on the body, some also have head spines, whereas others are devoid of spines (Fig. 2); caterpillars are entirely herbivorous, and the particular groups exhibit strong associations with particular plant families; the pupae are typically suspended.

Subfamily Charaxinae: a group of 350–400 mainly tropical species; adults fly very fast with robust bodies; underside of wings typically camouflaged and leaf-like, some with brilliantly colored upperside (*Agrias, Prepona*, and *Charaxes*); all are palatable, with few Batesian mimics (*Consul* and *Euxanthe*); adults feed primarily on juices of rotting fruit, dung, and/or carrion (rarely flower nectar), and most inhabit the forest canopy; caterpillars have smooth bodies, often bearing a corona of head spines or projections; host plant families include Annonaceae, Celastraceae, Convolvulaceae, Euphorbiaceae, Fabaceae, Flacourtiaceae, Lauraceae, Myrtaceae, Piperaceae, Poaceae, Rhamnaceae, Rutaceae, Santalaceae, and Sapindaceae.

Subfamily Apaturinae: a small group of medium to large species that are mainly tropical (often placed within the Nymphalinae); males often with brilliant iridescence on wing upperside, the Neotropical species (*Doxocopa*) with proboscis and forelegs lime-green; adults are strong fliers and feed entirely on carrion and putrefying fruits; caterpillars are herbivores on Ulmaceae and have a smooth body, bifid tail, and head with a pair of strong horns.

Subfamily Satyrinae (satyrs, wood nymphs, and browns): a cosmopolitan group of 1500–2000 mainly tropical species that are generally dull brown (some notable exceptions) with well-developed eyespots on the wings; most feed on juices of rotting fruits, some on fungi, and some on flower nectar in temperate regions; all are palatable, with only few clear examples of Batesian mimicry from Asia (*Penthema, Zethera*, and *Elymnias*); caterpillars are smooth with bifid tails, and some bear paired head projections; host plant families include Arecaceae, Araceae, Cyperaceae, Heliconiaceae, Poaceae, and Selaginellaceae.

Subfamily Brassolinae (owls): 70–80 entirely Neotropical species ranging from medium to some of the largest butterflies known (*Caligo*) that fly at dawn and dusk, most with characteristic, large eyespots on the hindwing underside, and they are common in butterfly conservatories; most species feed entirely on rotting fruit juices, but a few with a strongly reduced proboscis (*Brassolis* and *Dynastor*) may not feed; caterpillars have smooth bodies, often with dorsal pseudospines, bifid tails, and multiple horn-like projections on the head; host plant families include Arecaceae, Bromeliaceae, Heliconiaceae, Musaceae, and Poaceae, and they can be pests in banana and palm plantations.

Subfamily Amathusiinae (fauns and duffers): an Indo-Australian group of approximately 80 species (sometimes placed in Brassolinae or Morphinae); medium to large butterflies that fly at dawn and dusk and feed on rotting fruit juices; the group is palatable (except for perhaps *Taenaris*) and shows little or no mimicry; caterpillars have smooth bodies, some with long, downy setae, bifid tails, often with paired head horns; host plant families include Arecaceae, Musaceae, Poaceae, and Smilacaceae.

Subfamily Morphinae (morphos; Fig. 2): a group of 40–50 Neotropical species of medium to very large butterflies; the large reflective blue species (*Morpho*) are immediately noticeably in nature and in butterfly conservatories, but others that fly in the forest understory (*Antirrhea* and *Caerois*) are seldom observed; all feed on rotting fruits, none are considered to be unpalatable, and none are mimics; a favorite of collectors and butterfly conservatories, surprisingly little is known of their natural history; caterpillars are covered with red, yellow, and green patterns, bear tufts of dorsal and lateral setae, possess bifid tails, and bear short, paired projections on the head (Fig. 2); host plant families include Arecaceae, Bignoniaceae, Fabaceae, Menispermaceae, Poaceae, and Sapindaceae.

Subfamily Calinaginae: a group represented by two to five species in the genus *Calinaga* restricted to the Himalayan regions; it is apparent that *Calinaga* forms mimicry complexes with *Parantica* (Danainae), but it is not clear if it is Batesian or Müllerian mimicry; caterpillars are smooth with short bifid tails and stout head horns and feed on Moraceae.

Subfamily Nymphalinae: a diverse, cosmopolitan group of more than 3000 species (sometimes split into Limenitinae and Nymphalinae) containing a tremendous range of size and color patterns; some are migratory (*Vanessa, Eunica*, and *Sallya*), some are unpalatable models, some are palatable mimics, some show tremendous seasonal polymorphism (*Junonia*), some feed on flower nectar, others feed on rotting fruits and carrion; some groups inhabit tropical forest canopies (*Euphaedra, Cymothoe, Panacea, Epiphile*, and *Baeotus*), some pass north temperate zone winters as adults (*Nymphalis*); caterpillars usually covered in spines, many with well-developed pairs of head spines (Fig. 2); pupa often with bifid projections on head; host plant families include Acanthaceae, Caprifoliaceae, Convolvulaceae, Euphorbiaceae, Fagaceae, Flacourtiaceae, Lamiaceae, Loranthaceae, Moraceae, Plantaginaceae, Poaceae, Rubiaceae, Rutaceae, Salicaceae, Sapindaceae, Scrophulariaceae, Urticaceae, and Verbenaceae.

Subfamily Acraeinae: a group of approximately 250 tropical African species, with a few in Asia and the neotropics; ranging from small to large, all are slow-flying unpalatable species that contain cyanogenic compounds, forming aposematic models for a variety of other groups, and are involved in Müllerian mimicry (*Acraea, Bematistes*, and *Actinote*); one species (*Acraea encedon*) is almost entirely female and reproduces via parthenogenesis; caterpillars (many which feed gregariously in communal nests) are densely covered in spines, but lack head spines; host plant families include Asteraceae, Passifloraceae, Sterculiaceae, Tiliaceae, and Urticaceae.

Subfamily Heliconiinae (passionflower butterflies): a group of 70–80 mainly Neotropical species with a few in Asia; most with well-developed eyes that are wider than the thorax, elongate forewings, and most serve as models in Batesian and Müllerian mimicry complexes; they feed on flower nectar, and some on pollen (*Heliconius* and *Laparus*) from which they can derive some of their unpalatability; caterpillars are densely spiny, bear paired head spines, and feed entirely upon plants in the Passifloraceae and allies—hence the name passionflower butterflies.

Subfamily Danainae (milkweed butterflies, tigers, and crows): a cosmopolitan group of approximately 150 species, with most in tropical Africa and Asia; ranging from medium to large size, most are slow flying, conspicuously colored, and models in mimicry complexes (*Danaus, Amauris, Idea, Tirumala*, and *Parantica*); their distasteful nature derives from feeding as caterpillars on milkweeds, but they also acquire other chemical defenses by feeding as adults on flowers and wounds in certain plants containing particular alkaloids; males often have well-developed, brush-like scent hairs they use during courtship; caterpillars are smooth, often patterned with alternating bands of black, white, and yellow, and many have dorsal pairs of fleshy tubercles; the pupae are frequently reflective gold or silver; host plant families include Apocynaceae, Asclepidaceae, and Moraceae.

Subfamily Ithomiinae (ithomiines and glass wings; Fig. 1): approximately 300 entirely Neotropical species; adults typically with a very small head, elongate wings, and species vary in color from transparent (*Greta, Ithomia*, and *Pteronymia*) to bright tiger-striped patterns [*Mechanitis, Tithorea*, and *Melinaea* (see Fig. 1)]; they are involved in both Müllerian and Batesian mimicry, representing unpalatable models for many other groups; their unpalatable properties derive from larval host plants and chemicals acquired by adult feeding, and males typically possess a tuft of scent hairs between the wings that disseminate pheromones; caterpillars are smooth, often with fleshy tubercles, and may be brightly colored or cryptic; The pupa is often reflective gold or silver and squat; host plant families include Apocynaceae, Gesneriaceae, and Solanaceae.

Subfamily Tellervinae: a group of 6–10 Australasian species all in the genus *Tellervo;* they serve as a models for nymphaline and satyrine Batesian mimics; this group has been included within the Ithomiinae; the caterpillars resemble some Danainae and feed on Apocynaceae.

Subfamily Libytheinae (beaks and snout butterflies): a cosmopolitan group of approximately 10 species recognized by the large erect palpi that form a "snout"; most species are well-known to periodically undergo spectacular migrations; the caterpillars somewhat resemble those of the Pieridae, and all feed on Ulmaceae.

II. EARLY STAGES AND HOST RELATIONSHIPS

Like all members of Lepidoptera, butterflies have four discrete stages in the life cycle (egg, caterpillar, pupa, and the adult), each with particular characteristics, behaviors, and requirements. Furthermore, to complete their life cycle, butterflies require a plant or insect host to feed on.

A. Egg

Butterfly eggs are laid either singly or in small to large clusters, either on or off the host, and the location where the egg is laid is typically important (Fig. 2). The eggshell frequently has an elaborate sculpturing that plays a role in respiration, and each major group of butterfly has its own form of egg.

B. Caterpillars

Depending on the group, the appearance of butterfly caterpillars may range from cryptic to aposematic, they may be covered with spines and/or hairs or appear to be naked, and their diets may include plant tissue or the flesh of other insects or they may feed entirely on secretions produced by other insects (Fig. 2). All caterpillars consist of three major sections: the head, thorax, and abdomen. The hardened head houses mandibles that function to shear off bites from their food. The head bears a gland that lays down silk that is grasped as the caterpillar moves forward, helping the caterpillar grip the substrate and also to secure rolled leaves in which they may shelter. Each thoracic segment bears a pair of true legs, whereas the 10 abdominal segments form the bulk of the body, housing the long gut. Caterpillars walk by using the prolegs (segments 3–6 and segment 10), and these function by hydraulic pressure and muscles. Segments 1–8 bear the external ring-like orifices (spiracles) that allow gas exchange with the atmosphere.

After reaching a particular size each caterpillar instar stops eating and undergoes a molt to the next instar; this is how caterpillars grow. Generally, but not always, there are five larval instars followed by a molt to the pupa, or chrysalis. Caterpillar growth is not a steady increase in weight from first to final instar; rather, there is a dramatic fluctuation of weight between each molt. Newly molted caterpillars may weigh about the same as or even less than previous instars, but their weight will quickly increase and exceed that of previous instars.

Caterpillar feeding behavior often differs among instars. In many groups, late-instar caterpillars may stop feeding during the day and then feed entirely at night. Alternatively, some lycaenid caterpillars may start life feeding on plants and, after molting to a later instar, they fall off the plant, are picked up by ants, and are carried into their nest where the caterpillars feed as carnivores on ant brood.

C. Pupa

When the final-instar caterpillar is fully grown, it stops eating and enters its molt to the pupa, or chrysalis. Within the pupa the caterpillar tissues are reconstructed into the adult by the process known as metamorphosis. The pupa attaches to a substrate by a series of hooks on the last segment (cremaster), and major groups typically have characteristic manners of pupation. For example, pupating with the head downward attached only by the cremaster is typical of Nymphalidae, but pupae of Papilionidae and some Pieridae attach by the cremaster with head upward and are restrained by a silk girdle. Pupae of Lycaenidae produce a whirring, clicking, or buzzing sound with a rasp and file system on the abdomen that may be a defense against predators. Sound production is also known in some nymphalid pupae.

When mature, the pupa splits along its dorsal surface, and the adult ecloses. After eclosion, the adult normally hangs from the pupal shell or nearby with the wings suspended downward so that it can expand and dry its wings. If dried in a crumpled manner the wings are useless for flying, and the butterfly is effectively dead. As a rule, female butterflies are mated soon after or even before eclosion, exemplifying one the most potent laws of evolution—nature abhors a virgin.

D. Adult

The adult butterfly (Fig. 1) is incapable of additional growth but is capable of flight, mating, and reproduction. Like all insects, the butterfly body is composed of the head, thorax, and abdomen.

The most obvious features of the butterfly head are the large compound eyes composed of numerous facets (ommatidia) that cannot focus but are sensitive to movement, light, and certain colors. A pair of distally thickened antennae arise from between the eyes that vary in shape according to the group. The antennae function as sensory organs for finding food, mating, and balance during flight and are sensitive to volatile chemicals. Between the eyes there is a pair of appendages called labial palpi, and between them lie the proboscis, a hollow tube composed of two interlocking halves that is coiled like a watch spring when not in use and can be extended for feeding. By virtue of having a proboscis, butterflies are restricted to a liquid diet that may include flower nectar, the juices of rotting fruit, carrion, dung, or semidigested pollen. Proboscis length may vary according to the group; in some species it is nearly vestigial, thus precluding feeding as adults (*Brassolis* and *Liphyra*), whereas in others the proboscis measures more than 1.5 times the length of the body (*Eurybia*), allowing them to take nectar from a wide range of flowers.

The thorax is composed of three fused segments bearing the wings and legs, and it contains the muscles for locomotion and various internal organs. As in all insects, the adult butterfly has six legs, one pair per segment. Butterflies have four wings (two forewings and two hindwings) typically covered in scales that give butterflies their characteristic colors and patterns. The color patterns of butterflies result mainly from the covering of scales that are arranged like overlapping roof tiles. There are three notable types of scales. The pigmentary scales are colored by the deposition of melanin, pterins, or other chemicals. Structural scales generate blue, violet, copper, or green colors by reflecting particular wave lengths of incident light. Androconial scales store and disseminate chemical odors (pheromones) that are used in mating, and some of these scales may be physically transferred to the female during mating.

The abdomen houses the digestive and reproductive tracts, terminating in the reproductive organs (genitalia). Except for segments housing the genitalia, the abdomen can stretch when the gut becomes filled with food, and abdominal distention may be considerable in groups feeding on rotting fruits (Charaxinae). The penultimate abdominal segment of males bears two appendages (claspers) that open to expose the aedeagus (penis) and serve to grip the female's abdomen during mating. The female abdomen terminates in three openings: the anus, egg pore, and copulatory pore. The configuration of genitalia is used extensively in butterfly taxonomy.

E. Host Relationships

An important aspect in the butterfly life cycle is the ability of ovipositing females to find, and caterpillars to feed on, particular plants. The liaison with plants is so strong that many groups of butterflies only associate with particular taxonomic groups of plants; other plants are unacceptable to both caterpillars and ovipositing females. In the Lycaenidae this association may extend to particular species of ants or Homoptera. For example, caterpillars of milkweed butterflies (Danainae) only feed on plant families containing milky latex, and those of the Heliconiinae use plants in the Passifloraceae. Such patterns of host association in butterflies gave rise to Ehrlich and Raven's classic paper that developed the concept of coevolution. On the whole, host association records for the Papilionidae, Pieridae, and Nymphalidae are much more complete than for the Lycaenidae.

III. BUTTERFLY–ANT SYMBIOSES

The ability to form symbiotic associations with ants (myrmecophily) occurs only within the Lycaenidae. Here, caterpillars provide food secretions to ants in exchange for protection against insect predators such as social and parasitic wasps. To form these symbioses, caterpillars have a suite of unique adaptations that may include organs (collectively known as ant-organs) to produce food secretions, volatile chemicals, and sound, all of which work in concert to modify ant behavior and enhance the protective attitude of ants toward caterpillars. Recent studies on ant-organs indicate that myrmecophily evolved at least twice in the butterflies: once in the Riodininae and independently in other lycaenid subfamilies.

The widespread trait of myrmecophily within the Lycaenidae and the fact that lycaenids account for approximately 50% of all butterfly species led Pierce (1984) to suggest that myrmecophily has amplified speciation rates in this group. Indeed, the diversity of life histories in myrmecophilous butterflies can be exceedingly complex, encompassing herbivores, carnivores, and those that feed as caterpillars only on secretions, and the associations with ants range from mutualistic to completely parasitic or predatory.

A. Food Secretions

Ants pay close attention to particular abdominal segments bearing ant-organs that produce food secretions, which in some species are known to have high concentrations of amino acids and sugars. In some Riodininae, these consist of a pair of organs (tentacle nectary organs) on the eighth abdominal segment that can be extruded individually or simultaneously. In all other lycaenid subfamilies, this organ is a single dorsal pore on segment 7 (dorsal nectary organ). Ants are so intent on obtaining the secretions that they constantly antennate the caterpillar to solicit more, and in many cases, this is a good example of a general rule among participants in symbiotic associations— "you scratch my back and I'll scratch yours."

B. Semiochemical Production

Myrmecophilous caterpillars may have extrusible glands that seem to produce volatile chemicals (semiochemicals or pheromones) that alter the behaviors of attending ants. Some Riodininae have a pair of extrusible glands on the third thoracic (anterior tentacle organs), whereas other lycaenids have a pair of glands on the eighth abdominal segment (tentacle organs). In both cases, when extruded from the body these organs do not produce a liquid secretion but rather the tip is modified with spines that gives the appearance of a tiny feather duster. These spines likely provide a larger surface area to disseminate volatile chemicals that may be similar to ant alarm pheromones. Instead of anterior tentacle organs, a small group of Riodininae caterpillars (*Theope*) possess a corona of inflated setae around the head that appear to disseminate semiochemicals.

C. Call Production

The idea that caterpillars produce acoustic calls might seem unlikely. We now know, however, that myrmecophilous caterpillars produce substrate-borne calls that function in the formation and enhancement of their symbioses with ants, and these calls bear similarities to those produced by ants for communicating among themselves. In most Riodininae caterpillars calls are produced by a pair of mobile, grooved, rod-like appendages (vibratory papillae) arising from the distal edge of the first thoracic segment. An acoustical signal is produced when the grooves on the vibratory papillae grate against head granulations (Fig. 2). In other lycaenid caterpillars the call is produced by thickened bumps located ventrally between abdominal segments. Most concepts of insect communication suggest that acoustical calls evolved in a sexual context. However, caterpillar calls provide an example showing that, by forming symbiotic associations, the call of one species may evolve to attract another, unrelated species.

D. Ants and Caterpillar Associations

In general, myrmecophily in caterpillars occurs with a particular type of ant. A basic element among myrmecophilous caterpillars is that they typically form symbioses only with ant species that depend heavily on secretions as food—ants that also form symbioses with Homoptera and plants. Therefore, secretion-harvesting ants likely played a key role in the evolution of myrmecophily, whereas those ants that are predators or herbivores did not. An ecological consequence of the evolution with secretion-harvesting ants is that in any suitable contemporary habitat, a suite of caterpillar, plant, and Homoptera species all share the same species of ant symbionts.

Among myrmecophilous caterpillars there are two main categories of ant association. The most widespread category comprises an association in which a particular caterpillar species may be tended by a suite of secretion-harvesting ant species. The other category comprises an association in which a caterpillar has an obligate association with a single species of ant. In this case, female butterflies may require the presence of a particular ant species to lay their eggs since caterpillars are unable to form symbioses with any other ant species. Furthermore, it is in these types of associations in which some caterpillars are adopted by the ants, taken into the nest, and become parasites or predators of their hosts.

IV. BUTTERFLY MIMICRY AND DIVERSITY

When several species of butterflies share conspicuously bright color patterns and fly together in the same habitat, this is called mimicry—a widespread and important antipredator defense. Fundamentally, mimicry occurs when one species closely resembles another species, and based on outward appearance one or both are avoided by predators. Virtually all major taxonomic groups of butterflies exhibit mimicry, but the warningly colored (aposematic) models are found predominately in particular groups of Papilionidae, Pieridae, and Nymphalidae, whose caterpillars feed on poisonous plants. Mimicry is a complex and subtle topic, but there are two basic types of mimicry in butterflies, both involving species that advertise conspicuous color pattern to predators: Batesian and Müllerian mimicry.

A. Batesian Mimicry

This is the phenomenon whereby one or more palatable species resemble one or more unpalatable model species, and the palatable species gain protection by duping predators. Here, predators avoid mimics because they resemble unpalatable models. The best known basic example of Batesian mimicry involves the palatable viceroy (*Limenitis archippus*) and the unpalatable monarch butterfly (*Danaus plexippus*) in North America. However, Batesian mimicry occurs in all biogeographic regions, and one of the most intriguing examples involves the sex-limited mimicry of the wide-ranging African swallowtail, *Papilio dardanus*. Here, only the females are mimics. In this case, females are polymorphic with respect to their appearance, and their color patterns change geographically when flying with different model species. In other words, *P. dardanus* females from Kenya may look entirely different from those from Zaire or elsewhere, and in each instance they precisely mimic a different unpalatable model species. Thus, the type of female color pattern evolves in response to and depends on the local community diversity of the model species.

B. Müllerian Mimicry

This is the phenomenon whereby several unpalatable species (co-mimics) fly in the same area and share the same color pattern. In this case, the association of similarly patterned unpalatable species is thought to reinforce the nasty experience predators have when eating butterflies of such a color pattern, thus educating predators to avoid it.

C. Mimicry and Diversity

Aposematic species are an important element of butterfly diversity, particularly in the tropics. For example, approximately 20–25% of all Papilionidae, Pieridae, and Nymphalidae in Costa Rica and Kenya are clearly aposematic models. Although accounting for a large percentage of all butterfly species, there are few documented examples of aposematic Lycaenidae.

Müllerian mimicry is primarily a tropical phenomenon and is an important part of tropical faunal diversity. This can be appreciated by considering the diversity of color patterns shown by many co-mimetic species and races of ithomiine and heliconiine butterflies (plus many co-mimic moths and other butterfly groups) that converge across large areas of the neotropics. Even when the tightly overlapping distribution of only two co-mimetic species of *Heliconius* and their many precisely convergent color patterns are mapped over the Neotropical region, one cannot doubt the power mimicry plays in the evolution, organization, and diversification of tropical butterfly faunas. The numerous co-mimetic species of Acraeinae butterflies throughout the African region also serve as another potent example of the influence of mimicry on the diversity of butterfly communities. It is clear that butterfly mimicry evolved in the context of multispecies interactions which involved butterflies, their predators, and the plant and insect hosts they feed on as adults and caterpillars. Thus, mimicry serves to remind us that habitat disturbance and/or destruction may have important implications for the continued survival and future evolution of interacting species.

V. GEOGRAPHICAL PATTERNS OF BUTTERFLY DIVERSITY

One of the most frequently asked questions about butterflies is the following: How many are there? The answer is thought to be between 12,900 and 15,819 species. This is hardly a satisfying answer since the question has little biological context. In contrast, questions framed by comparing different geographical areas impart a sense of dynamics to butterfly diversity. For example, one might ask: Does butterfly species diversity change with respect to latitude? or How do different areas of Africa compare with respect to numbers of butterfly species? Indeed, comparing species numbers among areas shows that latitude, biogeographical region, area size, and taxonomic affinity all contribute to global butterfly species diversity. A caveat is in order about comparative species numbers: Sampling effort is almost always unequal among different areas, and the taxonomic accuracy of species counts or lists often varies among sites. Thus, it is prudent to use care when interpreting the generality of species diversity patterns among areas.

A. Latitudinal Gradients of Species Diversity

As is the case for most groups of terrestrial organisms, butterfly species diversity increases toward the equator. Comparing the North American and Neotropical faunas suggests that the neotropics has an estimated 10–12 times more butterfly species compared to North America (Table I). Comparing site diversity at different latitudes, however, provides a richer perspective of this phenomenon. A latitudinal transect using five well-known sites from the Americas makes it obvious that species numbers increase when moving from the north latitudes toward the equator (Table II), even though the smallest areas being compared are at the equator. The dramatic equatorial increase in species is exemplified by the fact that a mere 500 ha of lowland Ecuadorian forest has more butterfly species than all of North America.

The relative contribution of each butterfly family to site diversity also varies with latitude. From Table II, it is evident that the relative contribution of Papilionidae and Pieridae to total species diversity is about the same regardless of latitude, but the contributions of Nymphalidae and Lycaenidae increase dramatically between 20 and 0° latitude (Fig. 3). The increase in species

TABLE I

Approximate Numbers of American and World Butterfly Species

	North America	Neotropics	World
Papillionidae	35	169	621–645
Pieridae	63	347	1086–1105
Lycaenidae	171	2300–3000	6000–6900
Nymphalidae	202	1850–2500	5175–5975

TABLE II

Latitudinal Transect Showing Changes in American Butterfly Species Diversity

			Family				
Site	Latitude	Area	Papilionidae	Pieridae	Nymphalidae	Lycaenidae	Total species
H. J. Andrews Forest, Oregon	44°14′	6400 ha	5	7	27	23	62
Los Angeles Basin, California	34°4′	>1 million ha	6	15	26	32	79
Estacion Los Tuxlas Mexico	18°35′	700 ha	14	19	113	66	212
La Selva, Costa Rica	10°26′	1000 ha	16	26	219	181	442
Garza Cocha, Ecuador	0°29′	500 ha	24	25	314	312	676

of Nymphalidae is accounted for by subfamilies that reach their greatest diversity in the Amazon (Morphinae, Brassolinae, Charaxinae, Satyrinae, and Ithomiine), whereas the increased contribution of Lycaenidae is due mainly to the subfamily Riodininae, which is also an Amazonian group.

B. Variation among Neotropical Sites

Species diversity is not equal among Neotropical forest sites. Nine well-known sites that have many species in common vary strongly in numbers of species and in the contribution of each family to species diversity (Table III). Here, species numbers may range from 212 in Mexico to 1199 at one Brazilian site. Among six Amazonian sites, numbers range from 676 to 1199 species. When averaged over the entire Table III, the contribution of each family to species diversity is as follows: Papilionidae, 3.8%; Pieridae, 4.8%; Nymphalidae, 45.2%; and Lycaenidae, 46.1%. It may be significant that these proportions approximately reflect those observed when comparing latitude (Fig. 3).

Comparisons among sites from the Americas point to the general relationship of the higher species richness that occurs at lower latitudes and to the marked variation among sites that share many species in common. These examples also show that the relative contribution of each family to species richness is not equal; the proportion of nymphalid and lycaenid species increases toward the equator, but the proportion of papilionid and pierid species appears relatively constant. At this point, we might ask if there are similar patterns outside the neotropics.

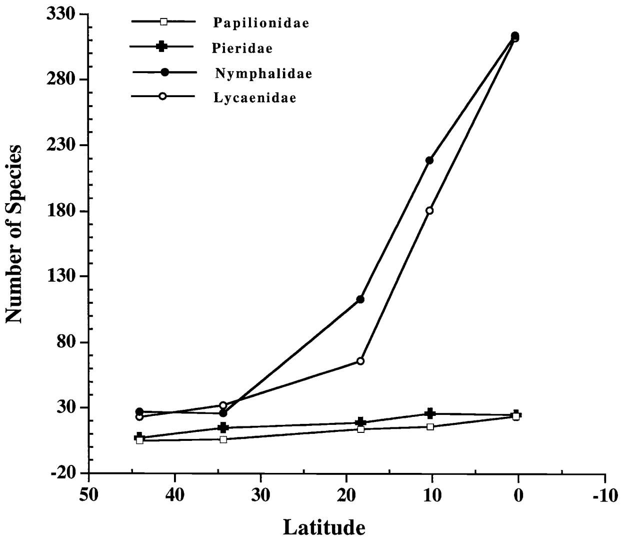


FIGURE 3 Latitudinal Gradient of American Butterfly Species Diversity. Each line represents species numbers per family contributed to five sites distributed from 44° north to the equator.

C. Regional Patterns of Species Diversity

Although most butterfly species are tropical, the number of species is not distributed equally among tropical regions (Table IV). As one might expect, those areas closest to the equator have a greater number of species, but there is also an effect of size on species diversity (Table IV). In this comparison, the continent of Africa has by far the most species, but it also encompasses the greatest geographical area and habitat types (deserts, mountains, forests, and savannas) of any area. Within Africa there are more species in Zaire than in Kenya or Southern Africa, highlighting the ecological diversity in African climate and habitat types. Thus, the entire continent of Africa should contain more species than a geographic subset.

As in the neotropics, the relative contributions of families to total species numbers differ among areas of Africa (Table IV). Here, the Papilionidae and Nymphalidae contribute equally to total richness of both Kenya

TABLE III

Comparative Diversity among Neotropical Sites[a]

Site	Family				Total species	Total area (ha)
	Papilionidae	Pieridae	Nymphalidae	Lycaenidae		
Estacion Los Tuxlas Mexico	14 (6.6)	19 (8.9)	113 (53.3)	66 (31.1)	212	700
Finca La Selva, Costa Rica	16 (3.6)	26 (5.8)	219 (49.5)	181 (40.9)	442	1,000
Garza Cocha, Ecuador	24 (3.5)	25 (3.7)	315 (46.6)	312 (46.1)	676	500
Jatun Sacha, Ecuador	25 (3.6)	23 (3.2)	306 (43.7)	345 (49.3)	699	600
Paquitza, Manu, Peru	25 (2.9)	31 (3.6)	369 (43.3)	427 (50.1)	852	3,900
Tambopata, Peru	25 (3.1)	26 (3.3)	337 (42.3)	409 (51.3)	797	>2000
Cacaulandia, Brazil	30 (2.5)	31 (2.6)	423 (35.3)	715 (59.6)	1199	2,000
Alto Jurua, Brazil	38 (3.7)	37 (3.6)	467 (45.0)	496 (47.8)	1038	>500
Cerro do Japi, Brazil	19 (4.7)	36 (8.9)	193 (47.8)	156 (38.6)	404	>400

[a] The percentage of the total species at a particular site is in parenthesis.

and Zaire, but Kenya shows an increased proportion of Pieridae and a decreased proportion of Lycaenidae. On the other hand, the proportional contributions of Papilionidae and Pieridae are reversed in the Zaire fauna. The Papilionidae and Nymphalidae in the Southern African fauna contribute less to the total than other mainland African areas, but nearly 55% of all Southern African butterflies are Lycaenidae. Overall, these proportional differences (Table IV) may reflect the ecological responses of taxonomic groups to differences in habitat types and/or the differences among regional taxonomists. Finally, in comparison to mainland Africa the fauna of Madagascar is interesting. Here, only 17% of all butterflies are lycaenids, but nymphalids constitute 67% of the total fauna, most of which are Satyrinae (i.e., 42% of all the butterflies). This example demonstrates how historical colonization and subsequent radiation of an island by one group can produce a fauna distinctly different in composition from that of the mainland.

Only a broad-brush comparison among areas of different sizes and regions is necessary to appreciate that, overall, butterfly diversity is highest in the neotropics (Table I–Table V). Despite the great disparity in geo-

TABLE IV

Variation in the Contribution of Each Family to Species Richness among Different Faunas[a]

Fauna	Family				Total
	Papilionidae	Pieridae	Nymphalidae	Lycaenidae	
All of Africa	80 (2.9)	145 (5.3)	1107 (40.6)	1397 (51.2)	2729
Zaire	48 (3.7)	100 (7.6)	607 (46.5)	551 (42.2)	1306
Kenya	27 (3.7)	87 (12.1)	335 (46.5)	271 (37.6)	720
Southern Africa	17 (2.3)	54 (7.2)	265 (35.6)	409 (54.9)	745
Madagascar	13 (4.9)	28 (10.7)	175 (66.8)	46 (17.5)	262
Australia	18 (6.5)	35 (12.6)	85 (30.6)	140 (50.3)	278
New Guinea	41 (5.2)	146 (18.6)	222 (28.3)	376 (47.9)	785
Malaysia	44 (5.8)	44 (5.8)	273 (35.9)	400 (52.5)	761
Costa Rica	42 (4.0)	71 (6.8)	438 (41.9)	493 (47.2)	1044

[a] The percentage of total faunal richness by each family is given in parentheses.

TABLE V

Percentage of World Species by Geographical Region[a]

	Region						
	North America	Neotropics	Costa Rica	Africa	Australia	New Guinea	Maylasia
Total species	471	5341	1044	2729	278	785	761
% World fauna	3.4	38.8	7.6	19.8	2.0	5.7	5.5

[a] Total number of butterfly species for world fauna (13, 753.5) and the neotropical region (5341) represent averages.

graphical area in the comparison, the small country of Costa Rica has a greater percentage of the world's butterfly species than does North America, Australia, New Guinea, or the peninsula of Malaysia (Table V). Since our taxonomic understanding of African, Australian, and Oriental butterflies is more thorough than that of the neotropics, it is likely that the total Neotropical butterfly species count will increase in the future.

D. Species Diversity in Space and Time

When estimating species diversity in any habitat, the variables of time, space, and sample size can be profoundly important. For example, consider two samples taken in the same Wisconsin prairie—one for 7 days in mid-July and the other for 7 days in mid-January. Despite equal sampling effort, the number of species tallied in January would certainly be zero because no butterfly is known to fly in a prairie during midwinter. Here, it is easy to see that time of sampling is important for estimating species diversity.

The problem of sampling in space and time becomes more important in forest areas, especially tropical ones. Consider a recent study of fruit-feeding nymphalid butterflies conducted on 600 ha of lowland Ecuadorian rain forest that investigated how species were distributed in space and time. To ensure equal sampling, butterflies were simultaneously trapped in the canopy and understory for the first week of every month for a period of 1 year. The results showed that both time of sampling (month) and position of trap (canopy or understory) were extremely important in estimating species diversity at this site (Fig. 4). In other words, had the study sampled for only a few months, or only in the understory, species diversity would have been greatly underestimated. A similar study from a different site in Ecuador also demonstrated that a significant proportion of species diversity was accounted for by time of sampling and the forest canopy fauna. This study further showed that species diversity varied significantly over short distances within the same, seemingly uniform rain forest, and it highlights the importance of sample size in comparisons of species diversity (Fig. 5). Such studies illustrate the critical nature of sampling methods, space, time, and sample sizes in comparisons of butterfly diversity.

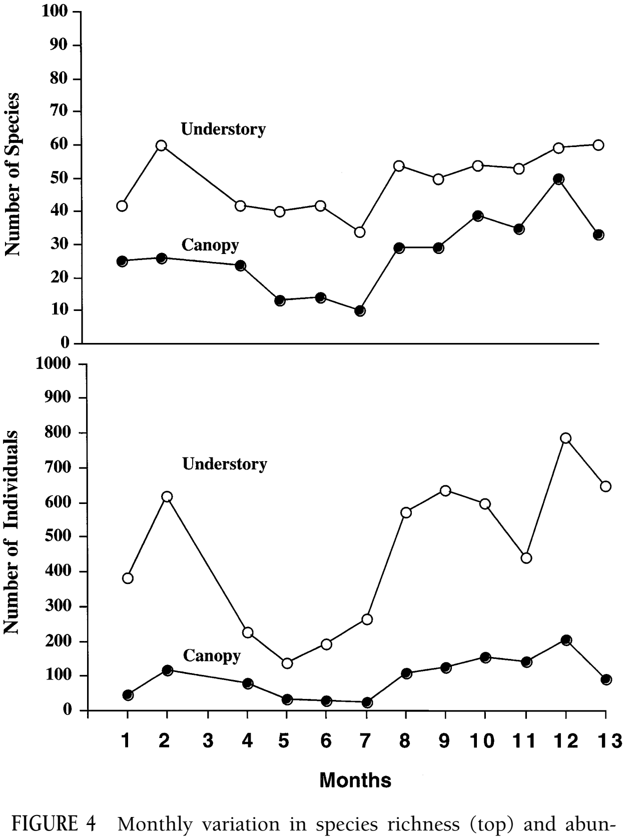


FIGURE 4 Monthly variation in species richness (top) and abundance (bottom) of fruit-feeding Nymphalidae in forest canopy and understory (from DeVries *et al.*, 1997).

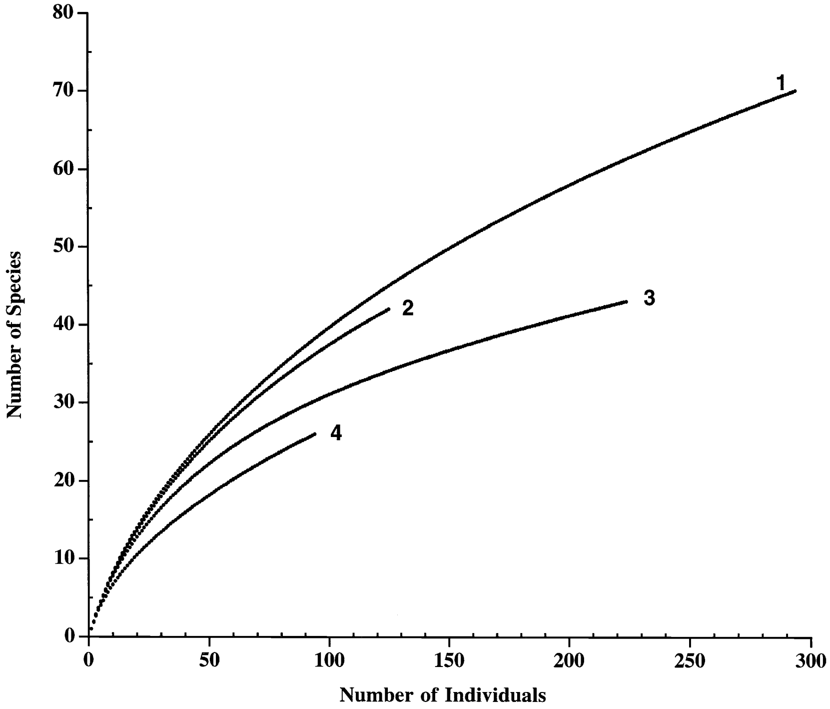


FIGURE 5 Species diversity of fruit-feeding Nymphalidae showing the relationship to space and sample-size. Each numbered line represents a plot sampled from within a single tract of continuous rainforest (from DeVries *et al.*, 1999).

VI. BUTTERFLY DIVERSITY AND HABITAT DESTRUCTION

It is obvious that the evolution of butterfly diversity is based on historical and contemporary interactions with many species. These biological interactions include plant and/or insect hosts, co-mimics in Batesian and Müllerian mimicry complexes, predators, and parasites. Butterflies have also evolved within and adapted to a great many biomes, habitats, and microhabitats, ranging from the multilevels within lush tropical rain forests to starkly dry deserts and subarctic tundra.

Habitat destruction always has profound effects on the biological communities that inhabit them, and butterflies are no exception. Like all organisms, butterflies live, evolve, and diversify within dynamic biological systems, and as such they cannot be studied as art objects or protected as inventoried stock. To date, butterflies have served as tools for understanding the diversification of life on Earth and the fundamental interactions among species. However, our future understanding of butterfly diversity will depend on a renewed interest in studying them in the natural world and valuing the habitats in which they occur.

Acknowledgments

The author is indebted to P. Ackery, G. Austin, G. Beccaloni, S. Collins, H. Greeney, A. Heath, R. Hill, R. Lande, T. Larsen, D. Lees, J. Mallet, N. Pierce, K. Rozema, T. Walla, and especially C. M. Penz for providing information, encouragement, assistance, and boundless enthusiasm about butterfly diversity.

See Also the Following Articles

HYMENOPTERA • INSECTS, OVERVIEW • MOTH • SPECIES COEXISTENCE

Bibliography

Ackery, P. R. (1984). Systematic and faunistic studies on butterflies. *Symp. R. Ent. Soc.* **11**, 1–21.

Ackery, P. R. (1988). Host plants and classification: A review of nymphalid butterflies. *Biol. J. Linnean Soc.* **33**, 95–203.

Ackery, P. R., and Vane-Wright, R. I. (1984). *Milkweed Butterflies: Their Cladistics and Biology*. British Museum (Natural History), London.

Beccaloni, G. W. (1997a). Vertical stratification of ithomiine butterfly (Nymphalidae: Ithomiinae) mimicry complexes: The relationship between adult flight height and larval host-plant height. *Biol. J. Linnean Soc.* **62**, 313–341.

Corbet, A. S., Pendlebury, H. M., and Eliot, J. N. (1992). *The Butterflies of the Malay Peninsula*. Maylayan Nature Society, Kuala Lumpur.

DeJong, R., Vane-Wright, R. I., and Ackery, P. R. (1996). The higher classification of butterflies (Lepidoptera): Problems and prospects. *Entomol. Scand.* **27**, 65–101.

DeVries, P. J. (1987). *The Butterflies of Costa Rica and Their Natural History. I. Papilionidae, Pieridae & Nymphalidae*. Princeton Univ. Press, Princeton, NJ.

DeVries, P. J. (1992). Singing caterpillars, ants and symbioses. *Sci. Am.* **267**, 76–82.

DeVries, P. J. (1997). *The Butterflies of Costa Rica and Their Natural History. II. Riodinidae*. Princeton Univ. Press, Princeton, NJ.

DeVries, P. J., Walla, T. R., and Greeney, H. (1999). Species diversity in spatial and temporal dimensions of fruit-feeding butterflies from two Ecuadorian rainforests. *Biol. J. Linnean Soc.* **68**, 333–353.

Ehrlich, P. R., and Raven, P. (1964). Butterflies and plants: A study in coevolution. *Evolution* **18**, 596–604.

Larsen, T. (1991). *Butterflies of Kenya and Their Natural History*. Oxford Univ. Press, Oxford.

Mallet, J. B. L., and Joron, M. (1999). Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation. *Annu. Rev. Ecol. Syst.* **30**, 201–233.

Neild, A. F. E. (1996). *The Butterflies of Venezuela. Part 1: Nymphalidae I*. Meridian, London.

Parsons, M. (1999). *The Butterflies of Papua New Guinea*. Academic Press, San Diego.

Pierce, N. E. (1987). The evolution and biogeography of associations between lycaenid butterflies and ants. *Oxford Surv. Evol. Biol.* **4**, 89–116.

Pierce, N. E. (1995). Predatory and parasitic Lepidoptera: Carnivores living on plants. *J. Lepidopterists Soc.* **49**, 412–453.

Pringle, E. L. L., Henning, G. A., and Ball, J. B. (Eds.) (1994). *Pennington's Butterflies of Southern Africa*. Struik Winchester, Capetown.

Scriber, J. M., Tsubaki, Y., and Lederhouse, R. C. (Eds.) (1995). *Swallowtail Butterflies: Their Ecology and Evolutionary Biology*. Scientific Publishers, Gainesville, FL.

Turner, J. R. G., and Speed, M. P. (1996). Learning and memory in mimicry. I. Simulations of laboratory experiments. *Philos. Trans. R. Soc. B* **251**, 1157–1170.

Tyler, H. A., Brow, K. S. Jr., and Wilson, K. H. (1994). *Swallowtail Butterflies of the Americas*. Scientific Publishers, Gainesville, FL.

Vane-Wright, R. I., and Ackery, P. R. (Eds.) (1984). The biology of butterflies. *Symp. R. Entomol. Soc.* **11**, 1–429.

C_4 PLANTS

Rowan F. Sage
University of Toronto

GLOSSARY

bundle sheath Cell layer at the periphery of the vascular bundles in leaves. In C_4 plants, the reactions of the photosynthetic carbon reduction cycle are localized in this layer, in contrast to C_3 plants, where they occur in the mesophyll tissue. In comparison to C_4 species, the bundle sheath cells of C_3 species are small, with few chloroplasts.

Kranz anatomy Specialized anatomy of the C_4 leaf in which the bundle sheath tissue is enlarged and enriched with chloroplasts, whereas the mesophyll is reduced in size and often forms a lighter green halo around the bundle sheath. This produces a wreath-like ("Kranz" in the original German) appearance.

mesophyll tissue Photosynthetic cells that are located between the arrays of vascular bundles and bundle sheath cells of a leaf. In C_4 plants, PEP carboxylase and pyruvate-phosphate dikinase are localized in the mesophyll tissue, while Rubisco is absent.

NUE Nitrogen-use efficiency, a measure of photosynthesis or growth relative to the nitrogen content of the leaves or plant.

phosphoenolpyruvate carboxylase (PEPCase) PEP carboxylase, the initial carboxylation enzyme in C_4 photosynthesis.

photorespiration Biochemical process in which O_2 is assimilated by the oxygenation of RuBP, and CO_2 is given off in the metabolism of the products of RuBP oxygenation.

Rubisco Ribulose-1,5-bisphosphate carboxylase/oxygenase, the enzyme in all plants that catalyzes the formation of two phosphoglycerate molecules from RuBP and CO_2, and catalyzes the competing reaction of RuBP oxygenation.

WUE Water-use efficiency, a measure of the amount of photosynthesis or growth per unit of water lost in transpiration.

IN PLANTS, PHOTOSYNTHESIS OCCURS as one of three biochemical syndromes. Most of the estimated 270,000 terrestrial plant species employ C_3 photosynthesis, in which CO_2 is directly assimilated (fixed) in a reaction catalyzed by ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) (Fig. 1). Between 20,000 and 30,000 species utilize Crassulacean acid

Encyclopedia of Biodiversity, Volume 1

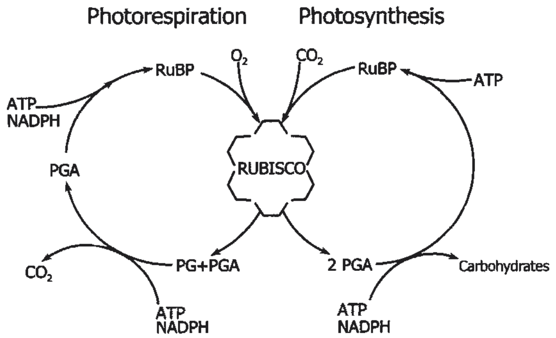


FIGURE 1 Diagrammatic representation of the photosynthetic carbon reduction cycle and the photorespiratory cycle in C_3 plants. Abbreviations: PG, phosphoglycolate; PGA, phosphoglyceric acid; RuBP, ribulose-1,5-bisphosphate. (From Sage *et al.*, 1999b.)

metabolism (CAM) photosynthesis, a syndrome in which nighttime fixation of CO_2 by phosphoenolpyruvate carboxylase (PEPCase) leads to the storage of carbon for refixation by Rubisco the following day. The third syndrome is C_4 photosynthesis, a CO_2 concentration mechanism also based on the initial carboxylation of phosphoenolpyruvate (PEP) by PEPCase. In C_4 plants, the C_4 biochemical pathway does not replace the C_3 pathway; instead it is a supplemental pathway that concentrates CO_2 in the tissue region where Rubisco and the C_3 photosynthetic biochemistry are localized. In so doing, inefficiencies associated with the C_3 pathway are largely avoided.

I. INTRODUCTION

Some 7600 advanced plant species have C_4 photosynthesis, making it the least used photosynthetic pathway in terms of species numbers. Despite representing less than 4% of all plant species, however, C_4 plants are ecologically significant in biomes that cover over 40% of the earth's landmass. They are well represented, if not dominant, in most tropical to warm-temperate grasslands, savannas, and salt marshes of the planet. C_4 plants are major contributors to agricultural productivity, particularly in tropical regions (Table I). Economically significant C_4 species are currently increasing in importance because of their high productivity in tropical regions where human population growth is the most rapid. In a negative light, C_4 grasses that respond well to human-caused disturbance are serious threats to biodiversity because they aggressively spread where disturbance of tropical forests is extensive. Many non-weedy C_4 species are themselves threatened because the landscapes where they occur are often suited for agricultural development. Future threats to C_4 biodiversity also arise from rising atmospheric CO_2 levels and terrestrial eutro-

TABLE I

Economically Significant C_4 Plants of the World[a]

Grasses	
Global crops	*Zea mays* (maize), *Sorghum bicolor* (sorghum), *Pennisetum glaucum* (pearl millet), *Saccharum officinalis* (sugarcane)
Regional cereals	*Eleusine coracana* (finger millet, Africa), *Setaria italica* (foxtail millets, N. Africa, Arabia), *Digitaria exilis, D. deflexa*, and *D. iburua* (fonios, Africa), *Eragrostis tef* (tef, Ethiopia), *Brachiaria deflexa* (animal fonio, Africa), *Coix lacrymi-jobi* (adlay, India), *Echinochloa colona* (Sawa, India), *Panicum sumatrense* (sama, India), *Setaria pumila* (korali, India), *Panicum sonorum* (sauwi, Mexico)
Forages	*Paspalum* (5 spp.); *Pennisetum* (2 spp.), *Panicum* (2 spp.), *Cenchrus ciliaris, Chloris guyana, Cynodon dactylon, Digitaria decumbens, Melinis minutiflora, Sorghum almum, Setaria anceps*
Weeds	*Cynodon dactylon* (bermuda grass), *Digitaria sanguinalis* (crabgrass), *Echinochloa crusgalli* (barnyard grass), *Echinochloa colonum, Eleusine indica, Imperata cylindrica, Sorghum halopense* (Johnson grass)
Turf grasses	*Cynodon dactylon* (bermuda grass), *Zoysia japonica* (zoysia), *Pennisetum clandestinum, Paspalum* spp.
Sedges (no major crops)	
Weeds	*Cyperus esculentus, Cyperus rotundus*
Other	*Cyperus papyrus* (papyrus, central Africa)
Dicots	
Minor crops	*Amaranthus edulis* (grain amaranth, Americas), *Amaranthus tricolor* (vegetable amaranth, Americas, India), *Portulaca oleracea* (purslane, S.E. Asia)
Weeds	Amaranthus (3 spp.), *Heliotriopium indicum, Portulaca oleracea, Chamaesyce (= Euphorbia) hirta, Salsola kali* (tumbleweed of cowboy folklore), *Tribulus terrestris* (caltrop)
Other	*Haloxylon* spp. (saxouls, 4 species provided fuel and fodder along the silk roads of central Asia)

[a] Modified from Brown (1999).

phication, phenomena that can enhance the performance of C_3 plants over their C_4 competitors.

Because C_4 plants are significant to humans and ecosystem processes, they are important in discussions of global biodiversity. In such discussions, however, C_4 species cannot be treated as simply another set of plants. Unique features of the C_4 biochemistry produce physiological and ecological characteristics that are distinct from those of the C_3 vegetation that might otherwise occupy a landscape. These distinct characteristics in turn affect processes at the ecosystem or landscape level in a manner that might not occur if C_4 species were absent from the landscape. To enhance understanding of the role of C_4 plants for global biodiversity, this article provides a background perspective on the function, geography, taxonomic diversity, and evolutionary origin of the world's C_4 biota.

II. PHYSIOLOGICAL SIGNIFICANCE

C_4 photosynthesis is often treated as a single biochemical pathway. This is an oversimplification. It is better regarded as a heterogeneous syndrome resulting from over 30 independent evolutionary modifications of C_3 leaf anatomy and biochemistry. Three distinct biochemical pathways and eight anatomical configurations have been described (Fig. 2). Together, these yield 14 recognizable patterns of C_4 photosynthesis. In addition, over a dozen species have biochemical characteristics intermediate between C_4 and C_3 photosynthesis. Given this variation, what general characteristics distinguish C_4 from C_3 plants?

A. Common Features

Four features are common to all C_4 species. First, PEP carboxylase activity is enhanced 5- to 50-fold over that of C_3 species, and all of it is restricted to photosynthetic mesophyll cells of the leaf and stem. PEPCase catalyzes the formation of oxaloacetic acid (OAA, a four-carbon or C4 acid) from a three-carbon acceptor molecule (PEP) and bicarbonate (HCO_3^-, formed when CO_2 dissolves in water) (Fig. 3). Second, Rubisco is localized in the bundle sheath cell layer that is generally located at the periphery of the vascular bundles of the leaf (see Fig. 2). Third, a decarboxylation reaction is catalyzed by one of three enzymes localized in the bundle sheath cells. The three decarboxylases are NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME), and PEP carboxykinase (PCK). Finally, the bundle sheath cells of C_4 plants are enlarged and have more chloroplasts when compared with C_3 species, while the mesophyll cells are reduced in number. In C_4 plants, one or two mesophyll cells are present for every bundle sheath cell; in C_3 species, more than four mesophyll cells are typically present for every bundle sheath cell.

CO_2 is concentrated in the bundle sheath when C_4 acids (malic or aspartic) derived from OAA diffuse from the mesophyll to the bundle sheath cells (see Fig. 3). There, the C_4 acids are decarboxylated to produce CO_2 and a three-carbon product (pyruvate or PEP). The CO_2 accumulates to very high levels, typically 5- to 10-fold over the atmospheric level, while the three-carbon residue diffuses back to the mesophyll cells where it will, if necessary, be converted to PEP with the expenditure of two ATP molecules per CO_2 fixed. The final CO_2 fixation step is the carboxylation of RuBP by Rubisco, using the CO_2 transported into the bundle sheath cells by the C_4 cycle. Metabolism of the resulting phosphoglycerate (PGA, the three-carbon or C3 product of RuBP carboxylation) into sugars occurs in the same reaction sequence as in C_3 plants. None of the enzymes in the C_4 cycle are novel, as similar biochemical pathways control stomatal aperture, cellular pH, and respiratory metabolism in C_3 species. The novel changes in C_4 evolution are altered patterns of gene expression that in turn affect tissue development, enzyme regulation, and the localization of enzymes within the leaf.

B. Photosynthetic Enhancements

1. Functional Role

The C_4 pathway has two immediate benefits: first, it raises the CO_2 content in the bundle sheath to levels enabling Rubisco to operate close to its maximum activity. Second, the build-up of CO_2 around Rubisco greatly reduces the inhibitory process of photorespiration. In photorespiration, Rubisco binds O_2 to RuBP to produce the two-carbon compound phosphoglycolate (PG) and PGA (see Fig. 1). While the PGA can be immediately metabolized back to RuBP by the carbon reduction cycle, PG has no apparent value to the plant and must be metabolized to avoid its accumulation to toxic levels. Photorespiration is the metabolic process in which PG is converted to PGA, but in this process 25% of the carbon in the pool of PG molecules is lost as CO_2. In addition, for every two RuBP oxygenation events, one ATP and the equivalent of one NADPH is required to recover the remaining 75% of the carbon in the PG pool. Because of the loss of previously fixed carbon and the additional energy required to scavenge carbon from PG, photorespiration is potentially a significant inhibition on photosynthesis.

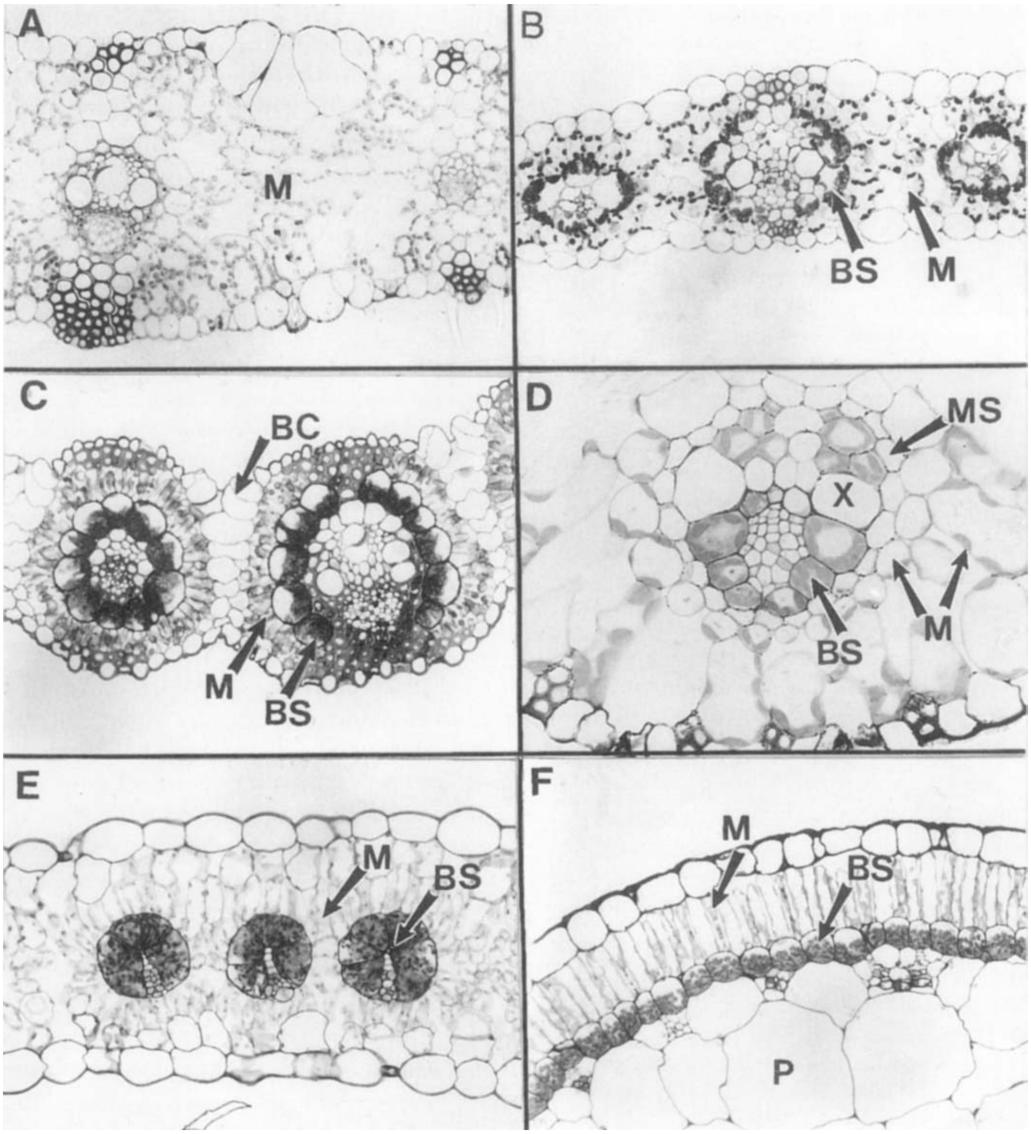


FIGURE 2 Transverse sections of C_3 and C_4 leaves, demonstrating five of the major anatomical patterns identified in leaves in C_4 plants. (A) The C_3 grass *Phalaris arundinacea*. (B–F) C_4 plants of different evolutionary origins: (B) *Zea mays*, a classic NADP-ME-type grass; (C) *Bouteloua gracilis*, a classic NAD-ME grass; (D) *Eleocharis vivipara*, an eleocharoid-type NAD-ME sedge; (E) *Atriplex rosea*, an atriplicoid-type NAD-ME dicot from the Chenopodiaceae; (F) *Salsola kamarovii*, a salsoloid-type NAD-ME dicot from the Chenopodiaceae. Arrows indicate mesophyll tissue (M), bundle sheath tissue (BS), and bulliform cells (BC). Xylem (X), mestome sheath (MS) and chloroplast-free parenchyma cells (P) are also indicated. (Photographs kindly supplied by professor Nancy Dengler, University of Toronto.)

There are three requirements for photorespiration to occur at a high rate (Sage, 1999). First, atmospheric O_2 must be abundant to provide the substrate for RuBP oxygenation. The current O_2 level of 21% is sufficient to support high levels of oxygenation if the other preconditions are met. Second, CO_2 levels in the atmosphere must be less than about 500 ppm (Fig. 4A). Elevated levels of atmospheric CO_2 (>1000 ppm) effectively inhibit the rate of RuBP oxygenation, while the current level of about 370 ppm allows for modest photorespiration at temperatures near 25°C. At present, the level of atmospheric CO_2 is anthropogenically enhanced by one-third over levels that persisted during the Holocene epoch of the past 10,000 years, and it is double the CO_2 level that predominated in the late Pleistocene (10,000 to 500,000 years ago). As a result of these lower CO_2 levels in recent geological time, photorespiration was higher in C_3 plants than it is today, and this likely

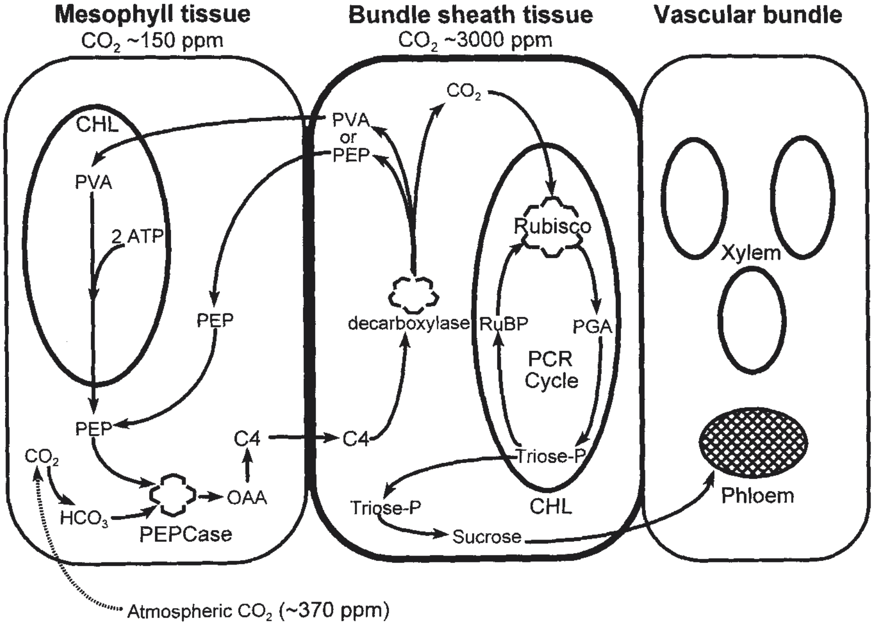


FIGURE 3 Diagrammatic representation of the C_4 photosynthetic cycle. The product of the PEP carboxylation reaction is oxaloacetic acid (OAA), which is converted to a transportable C_4 acid, either malic acid (NADP-ME type) or aspartic acid (NAD-ME and PCK). The decarboxylation reactions produce either pyruvic acid (PVA, produced by NADP-ME and NAD-ME) or PEP (produced by PEP carboxykinase). The PVA returns to the mesophyll chloroplasts, where it is converted to PEP using ATP. CHL, chloroplast; triose-P, triose phosphate; PCR, photosynthetic carbon reduction cycle or C_3 cycle.

reduced the photosynthetic potential of C_3 relative to C_4 vegetation.

The third requirement is warm temperatures. At low temperatures (<12°C), the rate of photorespiration is suppressed because Rubisco has low specificity for O_2 relative to CO_2. As temperature increases, the specificity of Rubisco for O_2 increases relative to that for CO_2 and the solubility of CO_2 relative to O_2 declines. Together,

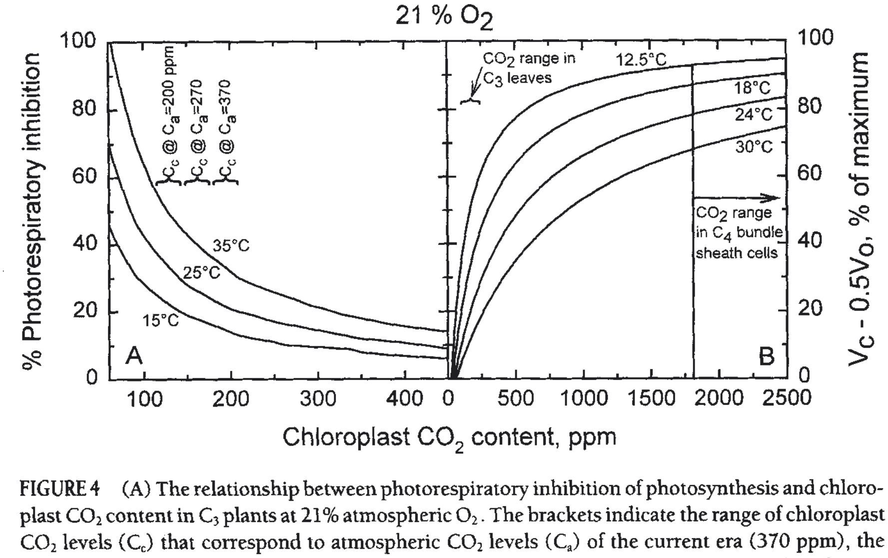


FIGURE 4 (A) The relationship between photorespiratory inhibition of photosynthesis and chloroplast CO_2 content in C_3 plants at 21% atmospheric O_2. The brackets indicate the range of chloroplast CO_2 levels (C_c) that correspond to atmospheric CO_2 levels (C_a) of the current era (370 ppm), the preindustrial Holocene era (270 ppm), and the late Pleistocene epoch (200 ppm). (B) The relationship between chloroplast CO_2 and carboxylation potential (V_c–$0.5V_O$) expressed relative to the maximum rate of Rubisco at CO_2 saturation. Responses modeled as described by Sage (1995).

these responses enhance RuBP oxygenation such that at 35°C, photorespiration inhibits photosynthesis by over 20% in current atmospheric conditions and by over 30% at preindustrial CO_2 levels of 270 ppm (see Fig. 4A). By contrast, because of the high level of CO_2 in the bundle sheath, photorespiration in C_4 plants is marginal (<5%) at all temperatures (Table II).

Temperature also affects the kinetics of Rubisco in a manner that has significance for C_4 photosynthesis (Fig. 4B). At low temperature (<12°C), Rubisco operates close to CO_2 saturation in the current atmosphere. As temperature increases, the amount of CO_2 required to saturate Rubisco rises, such that in C_3 plants in air, the enzyme is operating well below its maximum rate above 30°C. By raising CO_2 levels in the bundle sheath, C_4 plants are able to operate Rubisco near its maximum, substrate-saturated activity (V_{max}) at all temperatures (see Fig. 4B).

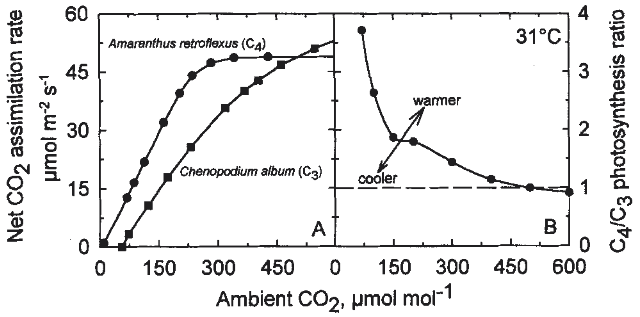


FIGURE 5 (A) The CO_2 response of photosynthesis in the weedy annuals *Amaranthus retroflexus* and *Chenopodium album* at 31°C. (B) The ratio of C_4 to C_3 photosynthesis for the responses in panel A. The double-headed arrow indicates the direction the curve would shift if measurements were conducted at different temperatures. (Modified from Sage, R. F., and Pearcy, R. W. (2000). The physiological ecology of C_4 photosynthesis. In: *Photosynthesis: Physiology and Metabolism* (R. C. Leegood, T. D. Sharkey, and S. von Caemmerer, Eds.), pp. 497–532. Kluwer Academic, Dordrecht, Netherlands, with permission.)

Together, the ability to reduce photorespiration and operate Rubisco near CO_2 saturation confers substantial photosynthetic advantages on C_4 species in warm climates and low-CO_2 atmospheres. These differences are best demonstrated by comparing the CO_2 response of photosynthesis of C_3 and C_4 species of similar ecological habitat, for example, prairie grasses or agricultural weeds (Fig. 5A). In such comparisons, the C_4 species always have CO_2 compensation points of photosynthesis near 0 ppm in contrast to their C_3 counterparts, for which the CO_2 compensation point is 40 to 60 ppm. C_4 plants have a steeper initial slope of CO_2 assimilation versus rising CO_2 than do C_3 species, which combined with their lower CO_2 compensation point enables them to have a higher photosynthesis rate at low CO_2 (Fig. 5B). C_4 photosynthesis becomes CO_2 saturated at CO_2 levels near the current atmospheric value, whereas C_3

TABLE II

Characteristics of C_3 and C_4 Crops and Weeds for Atmospheric Conditions of the Late Twentieth Century[a]

Attribute	C_3	C_4	Units
Maximum photosynthesis rate	20–50	35–75	μmol m^{-2} s^{-1}
Maximum daily production rate	10–40	40–80	g m^{-2} d^{-1}
Maximum biomass yield	1–5	3–8	kg DM m^{-2} yr^{-1}
Photosynthetic thermal optimum	15–30	30–40	°C
Photorespiration/photosynthesis @ 10°C	8%	1–5%	
@ 25°C	25%	1–5%	
@ 40°C	40%	1–5%	
$\delta^{13}C$	−20 to −35	−10 to −15	per mil
Intercellular CO_2	200–300	100–200	ppm
Maximum leaf nitrogen (N)	150–250	80–160	mmol m^{-2}
Relative Rubisco content	15–30	5–10	% of leaf nitrogen
Relative PEPCase content	<1	3–6	% of leaf nitrogen
Maximum efficiencies:			
Water use (WUE)	1.5–2.5	3–5	g DM kg^{-1} H_2O
Nitrogen use (NUE)	50–280	280–520	μmol CO_2 mol^{-1} N
Radiation use (RUE)	1.7–1.9	2.5	g DM MJ^{-1}

[a] DM indicates dry matter. Developed from Larcher (1995), Brown (1999), and Sage and Pearcy (1999).

TABLE III

Energy Requirements for Carbon Metabolism to Triose Phosphate, and Associated Quantum Yields of C_3 and C_4 Plants at 30°C[a]

	ATP	NADPH	Quantum yield at 30°C, 350 ppm CO_2
C_3 photosynthesis			
No photorespiration	3	2	0.078
With 32% photorespiration	5	3.2	0.053
C_4 photosynthesis			
NADP-ME and NAD-ME types			
C_3 cycle	3	2	
C_4 cycle	2	—	
25% leak rate	0.5	—	
3% photorespiration	0.15	0.1	
Total	5.65	2.1	0.055 to 0.068
PCK types (assuming 25% NAD-ME activity per unit of PCK activity)[b]			
C_3 cycle	3	2	
C_4 cycle	1.25	—	
25% leak rate	0.31	—	
3% photorespiration	0.15	0.1	
Total	4.71	2.1	0.06 to 0.65

[a] Adapted from Hatch (1987) and Kanai and Edwards (1999). The leak rate refers to fraction of pumped CO_2 that leaks out of the bundle sheath. The percentage photorespiration refers to the rate of photorespiration relative to photosynthesis.

[b] PCK-type plants require one ATP to power each PEP carboxykinase (PCK) reaction in the bundle sheath. This ATP is supplied by oxidation of malate in the bundle sheath mitochondria, with the malate coming from NAD-ME activity at a ratio of one NAD-ME event per three PCK events (Kanai and Edwards, 1999).

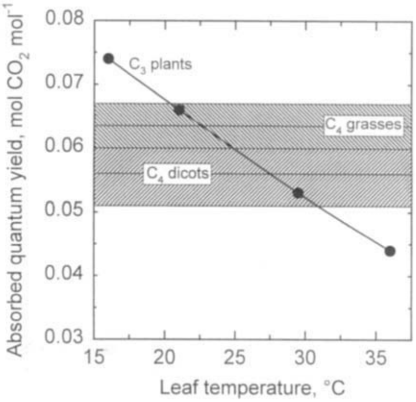


FIGURE 6 The quantum yield of C_3 and C_4 species as a function of leaf temperature at 350 ppm CO_2. In the C_3 response (represented by *Avena sativa*), rising temperature decreases quantum yields because of increased photorespiration. In the C_4 species, quantum yields are independent of temperature, and the hatched regions indicate the range of quantum yields observed in numerous C_4 dicot and grass species. (Developed from data in Ehleringer and Pearcy, 1983.)

plants show substantial photosynthetic enhancement with CO_2 enrichment above the current atmospheric value. Thus, at elevated CO_2, C_3 species typically have higher rates than their C_4 competitors in short-term gas exchange studies. In warm environments, these differences translate into significant photosynthetic and growth advantages for C_4 species in the atmospheric conditions of the late twentieth and early twenty-first centuries. In crops and weeds, for example, C_4 species on average have greater photosynthetic capacity, higher daily production rates, and yield more biomass during a growing season than do C_3 species (see Table II).

2. The Cost of C_4 Photosynthesis

In C_3 plants without photorespiration, each CO_2 costs 2 NADPH and 3 ATP to fix it into carbohydrate end products (Table III). In C_4 plants, 2 NADPH and 4.7 to 5.7 ATP are needed, assuming that 25% of the pumped CO_2 leaks out of the bundle sheath (Kanai and Edwards, 1999). The higher ATP cost of C_4 photosynthesis reflects the energy required to pump CO_2 into the bundle sheath. As the rate of photorespiration increases in C_3 plants, the energy costs of photosynthesis rise, such that they are equivalent to that of C_4 plants at 25° to 30°C, and greater above about 30°C (see Table III). Differences in the energy requirement of photosynthesis are demonstrated by comparing differences in the light-use efficiency (quantum yield) of C_3 and C_4 plants as a function of temperature (Table III and Fig. 6). C_3 species have superior quantum yields at cool temperatures, but lower quantum yield than C_4 species at warm temperature. Using quantum yields, it is possible to model the relative performance of C_3 versus C_4 photosynthesis as a function of CO_2 and temperature; as temperature increases, the CO_2 level at which C_4 species perform as well as C_3 plants increases (Fig. 7). For unexplained reasons, C_4 grasses have higher light-use efficiencies than do C_4 dicots, and thus begin to outperform C_3 species at cooler conditions and higher CO_2 levels.

3. Enhancement of Water- and Nitrogen-Use Efficiency

Second-order consequences arise as a result of the direct effects of the C_4 photosynthetic pathway on the catalytic efficiency of Rubisco. Because the C_4 cycle has a high capacity to concentrate CO_2 into the bundle sheath, even at low atmospheric CO_2 levels, C_4 plants will have less open stomata than C_3 species with the same photosynthetic rate. This reduces the rate of transpiration in C_4 relative to C_3 plants and leads to significant enhancement in water-use efficiency. Typically, C_4 species have two to three times the water-use efficiency of ecologically similar C_3 species (see Table II). Even at very low stomatal apertures that would yield negligible rates of CO_2 assimilation in C_3 plants, C_4 plants are able to have significant rates of photosynthesis. This is most important in hot, arid regions, where evaporative demands are so high that anything but very low stomatal apertures could lead to unsustainable rates of water loss. High water-use efficiency also assists C_4 species in saline soils, apparently by reducing the volume of transpirational water from which salt must be filtered.

In addition to improved water use, C_4 plants have higher nitrogen-use efficiencies of photosynthesis, typically 1.5 to 3 times that of C_3 species (see Table II). Because C_4 plants operate Rubisco near CO_2 saturation and greatly reduce oxygenase activity, they require about 25% as much Rubisco as C_3 species of equal photosynthetic capacity, and have very low levels of photorespiratory enzymes. As a consequence, C_4 plants have about 30% less nitrogen per unit leaf area as do ecologically similar C_3 species. These differences promote superior performance of C_4 grasses in nitrogen-deficient soils, such as those of the prairie grasslands of temperate North America.

III. C_4 BIOGEOGRAPHY

Clear patterns are evident in the global distribution of C_4 plants that are related to their distinct photosynthetic responses to temperature and light. As a result of their superior photosynthetic performance at high temperature, and their lower quantum yields below 25°C, C_4 species have superior growth in warmer but not cooler conditions, yet perform poorly in deep shade relative to similar C_3 species. The primary requirements for the success of C_4 plants are therefore warmer temperatures and access to at least moderate light intensities. Aridity, soil nitrogen supply, and salinity also affect the success of C_4 plants, but these are secondary in that they become important in warm, illuminated habitats.

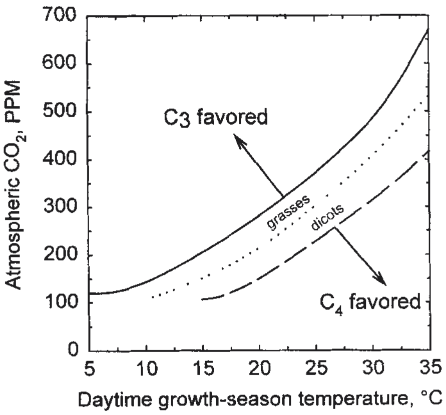


FIGURE 7 The effect of temperature on the CO_2 level at which the photosynthetic performance of C_3 species equals that of C_4 species. Responses were modeled using quantum yield differences between C_3 and C_4 species. (Adapted from Ehleringer *et al.*, 1997, with permission.)

A. Primary Controls: Temperature and Light

The critical temperature affecting C_4 plant distribution is growth season temperature. In temperate zones, C_4 plants tolerate severe cold outside of the growth season as well as co-occurring C_3 species, and often tolerate low night temperature as well as their C_3 associates. During the growing season, however, leaf temperature must routinely rise above 25°C for C_4 plants to be present in a community. Where daily temperatures commonly rise above 30°C during the growing season, they generally dominate grass and sedge floras of open landscapes. When plotted on the basis of growth season temperature (which reflects night and morning temperatures as well as the afternoon high), C_4 plants dominate grassland floras and biomass above 20–22°C. Above growth season averages of 23–24°C, C_3 grass productivity in open grasslands is very low (typically less than 10% of total biomass). At the other end of the spectrum, C_4 biomass is rare to absent where the temperature at the peak of the growth season averages less than 12–14°C. Some C_4 species occur at low frequency in cold climates, but these are usually restricted to warm microsites where daytime temperatures rise above the regional average.

The control of temperature over C_4 distribution is demonstrated by altitude and latitude transects that show the contribution of C_4 grasses to regional grass floras and grassland productivity. Where C_4 species dominate lowland floras, they decline in frequency as elevation increases, becoming uncommon above 2000 to 3000 m. Along latitude transects, the transition between C_3- and C_4-dominated floras corresponds to temperate latitudes, as is demonstrated by a survey of grass floras of the world's oceanic islands (Fig. 8). Above >50° latitude, C_4 species are rare or absent, and below temperate latitudes (<30°), C_4 grasses dominate the grass flora of any given island. On larger tropical islands, where extensive forests form, a significant number of C_3 forest grasses are present, so that the C_4 contribution

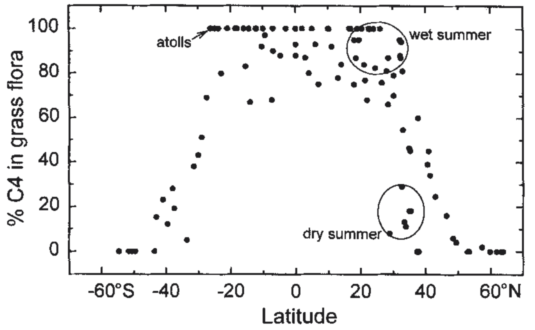


FIGURE 8 Percentage representation of C$_4$ photosynthesis in the grass flora of the oceanic islands of the world. The upper oval encloses islands from the western Atlantic/Pacific regions where summers are wet. The lower oval encloses islands from the eastern Atlantic/Pacific where summers are dry. (From Sage *et al.*, 1999b.)

may only be 70% of the grass flora. On smaller islands and atolls, where soils tend to be shallow, sandy, and drought-prone, grasses are often all C$_4$.

The second requirement for C$_4$ success is availability of moderate to high light intensity. Typically, this means that C$_4$ species require 30 to 50% of full sunlight intensities if they are to dominate a local habitat. In forest canopies where light intensities average less than 20% of full sun on a clear day, C$_4$ species rarely occur, regardless of latitude. Why C$_4$ species fail to dominate in low-light environments appears to result from a combination of factors. Shaded environments tend to be cool (<30°C at midday), so that high temperatures favoring C$_4$ species are not frequently encountered. In shade, much of the carbon that a plant acquires is during brief episodes of high light (sunflecks), and C$_4$ species appear to have a slightly reduced ability to exploit short sunflecks than do C$_3$ species. Also, because shaded environments tend to be cooler and more humid, C$_3$ plants often have more open stomates than they otherwise would. This allows for higher intercellular CO$_2$ levels in shade plants, thereby promoting greater photosynthetic efficiency.

The failure of C$_4$ species to do well in shaded settings has important consequences for the distribution of biomes around the globe. Except for a few desert shrubs that become arborescent with advanced age, and an unusual understory tree from Hawaii, there are no known C$_4$ trees or vines (Table IV), and relatively few C$_4$ herbaceous species survive in the interior of established forests. Thus, in warm to hot climates, the distribution of forest and woodland is the single greatest indicator of C$_4$ abundance and diversity. Where forests are present, the light environment will typically be too low to support a C$_4$ flora. Where forests are absent, either because of edaphic factors (e.g., shallow, nutrient-poor soils), severe abiotic stress (e.g., killing drought), mechanical stress (e.g., wind, waves, or flooding), logging, fire, or large animal activity, C$_4$ plants will be important components of the regional diversity. C$_4$-rich habitats include beaches along warm temperate to tropical coasts where wave action prevents woody dominance, tropical swamps and floodplains where flooding prevents dominance by a woody overstory, abandoned cropland at low latitudes, and warm temperate to tropical grasslands (Table V and Fig. 9). For example, in sub-Saharan Africa, the margins of lakes and slow-moving rivers are dominated by papyrus (*Cyperus papyrus*), a C$_4$ sedge. In the African savannas, elephants, giraffes, and fire aid in the reduction of C$_3$ *Acacia* trees that otherwise would dominate the landscape. Where elephants and fire become common, wooded landscapes are converted to C$_4$ grassland. Fire control and excessive poaching of elephants permit woodland formation, leading to a loss of the C$_4$-dominated grasslands.

B. Secondary Controls: Moisture, Nutrients, and Salinity

Aridity is often noted as directly supporting C$_4$ over C$_3$ biomass, because the higher WUE of C$_4$ species theoretically promotes their competitiveness in arid situations. In addition, interactions between aridity and other ecological factors, notably fire, are critically important. In the case of fire, aridity slows growth of woody species, thereby increasing their establishment time and the period during which they are vulnerable to fire injury. Aridity also enhances the probability of initial fire establishment, its subsequent rate of spread, and its intensity. In contrast to most woody species, C$_4$ grasses are typically fire tolerant, as their meristems are protected below ground or in fire-resistant tufts. Thus, the success of C$_4$ grasses in arid regions is promoted because fires typically kill or inhibit competing woody vegetation.

The WUE advantage of C$_4$ photosynthesis is perhaps most critical in very harsh locations, such as hot deserts in Africa, Asia, and northern Australia. Here, lack of rainfall combined with high daily temperature creates conditions where the transpiration potential may be greater than what a C$_3$ plant could support, particularly in the seedling stage. C$_4$ species not only survive, but they can thrive under thermal regimes that are harmful to most C$_3$ species. Consequently, C$_4$ plants are common if not dominant in the hot deserts of the earth. By contrast, the floras of deserts at high latitude or high elevation are exclusively C$_3$,

TABLE IV

Principal Life-forms of C_4 Plants[a]

Life-form	C_4 families	Examples
Trees	None	
Sub-trees (5 to 10)[b]	Chenopodiaceae	*Haloxylon* spp. (xerophytic halophytes and psammophytes)[c]
	Polygonaceae	*Calligonum* spp. (xerophytic psammophytes)
	Euphorbiaceae	*Chamaesyce forbesii* (small Hawaiian understory tree)
Vines	None	
Shrubs (500?)	Amaranthaceae	*Alternanthera, Aerva, Froelichia, Gomphrena, Guillemenia, Tidestromia* (all xerophytic)
	Chenopodiaceae	*Anabasis, Atriplex, Haloxylon, Salsola, Suada* (mostly xerophytic halophytes and psammophytes)
	Euphorbiaceae	*Chamaesyce* spp. (=*Euphorbia*, xerophytic shrubs)
	Polygonaceae	*Calligonum* spp. (xerophytic psammophytes)
Forbs (dicot herbs) (1000)	All dicot families	Summer desert annuals, xerophytic annual and perennial herbs of low latitudes, tropical and summer-active weeds, psammophytic herbs of low latitudes, tropical beach herbs
Graminoids (6000)	Poaceae	grasses, mainly of low-latitude and warm temperate summer habitats
	Cyperaceae	Sedges, mainly low-latitude and warm temperate wetlands

[a] Sub-trees refer to species that develop arborescence with age, but never rise above 12 m in height.

[b] Numbers in parenthesis refer to the estimated number of species in each life-form (after Sage *et al.*, 1999b).

[c] Psammophytes are plants of sandy soils.

TABLE V

A List of the Major Biomes of the World That Have High C_4 Plant Representation (Loosely Defined as >25% of the Vegetative Cover and a Potential for C_4 Dominance) or Have No C_4 Dominance and C_4 Cover Is Negligible (<1%)[a]

Biome	Major location
High C_4 representation	
Tropical, subtropical grassland and savanna	South and Central America, Africa, India, S.E. Asia, Australia
Warm temperate grassland and savanna	Central and S.E. North America, N. Argentina, Australia
Arid steppe (low to middle latitudes)	S. W. North America, central Asia, Australia, Africa
Beach dunes and bluffs, warm temperate to tropics	Global
Tropical, subtropical wetlands (nonarborescent)	Global, especially South America, central Africa, S.E. Asia
Salt marsh (warm temperate to tropical)	Global, but more temperate owing to mangrove dominance in tropics
Salt desert (<45° latitude)	W. North America, central Asia, central Australia
Hot deserts and semideserts	Southwest North America, Africa, Australia, S.E. Asia
Disturbed ground (low latitudes, low altitude)	Global, more in arid regions
Always C_3 dominated	
Forests (all, including arborescent wetlands)	Global
Tundra (all)	Polar latitudes, high mountains (all latitudes)
Heathlands	N. Europe, Canada, Russia
Cool temperate grasslands and savanna	Canada, N.W. U.S.A., S.E. Europe, S. Russia, Australia, S. Argentina, Tasmania
Montane grasslands	Global, elevations between 2000 and 4000 m
Mediterranean grasslands	California, Chile, S. Europe, Asia Minor, N. Africa, S. Africa, S.W. Australia
Mediterranean-type shrublands (chapparal)	California, Chile, S. Europe, Asia Minor, N. Africa, S. Africa, S.W. Australia
Temperate to boreal wetlands (nonarborescent)	Canada, N. Europe, Russia, New Zealand, Patagonia
Cold deserts and semideserts	W. North America, N. central Asia, Tibet
Salt marshes (>50° latitude)	Global
Mangles (mangrove swamps)	Tropical, sutropical latitudes
Disturbed ground (high latitude and altitude)	Global

[a] Biome list developed from *Ecosystems of the World* (1977–1990), D.W. Goodall, ed., Vols. 1–16, Elsevier, Amsterdam. Printed from Sage *et al.* (1999b).

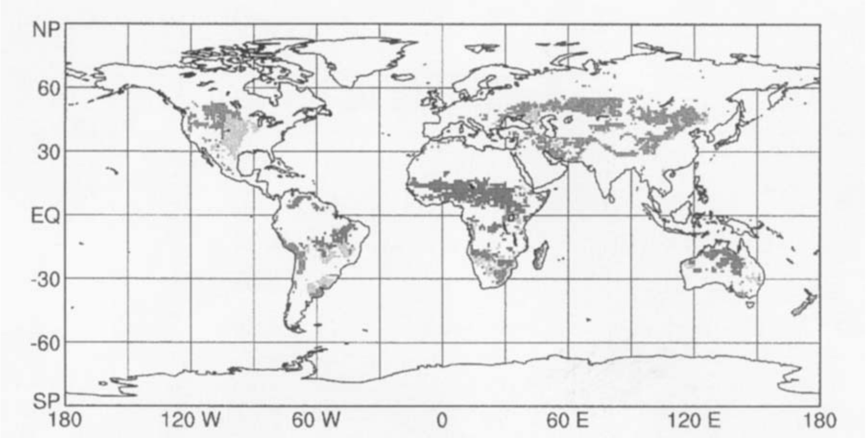


FIGURE 9 The distribution of the major grasslands of the world, according to Collatz *et al.* (1998). On the color insert, red areas indicate C_4-dominated grasslands, yellow areas are mixed C_3/C_4 grasslands, and blue areas are C_3-dominated grasslands. (Reprinted from Collatz *et al.*, 1998, with permission.)

reflecting the generally poor ability of C_4 species to compete in cold climates, regardless of any aridity that may be present.

The seasonality of precipitation is of great importance. Grassland ecosystems can be almost completely C_3 or C_4 depending on the timing of precipitation. Grasslands of Mediterranean climate zones that have cool-season precipitation (southern Europe, North Africa, California, Chile, southwestern Australia, and southwestern Africa) are heavily C_3 dominated, whereas at similar latitudes, temperate monsoon grasslands with significant summer precipitation are C_4 dominated. For example, the southwest tip of Africa near Capetown has a Mediterranean-type climate and <10% C_4 species in the grass flora. At Durban, on the eastern coast of South Africa at similar latitudes, the bulk of the precipitation falls during the warm season and the grass flora is 75% C_4 (Fig. 10). Oceanic islands show a similar pattern. In the eastern Atlantic and Pacific at temperate latitudes, summer rain is rare and C_4 presence is low (see Fig. 8). On the western edges of these oceans, summer precipitation is abundant as are C_4 grasses.

Seasonality of precipitation is particularly important in hyperarid regions of the world. In southwestern North America, for example, the Mojave Desert in California receives 60 to 90% of its precipitation during the winter, and as a consequence it has a high proportion (40%) of C_3 grasses in the grass flora. The Chihuahuan Desert in southwestern Texas receives the majority of its precipitation in summer, and C_3 grasses are infrequent (20% of the grass flora). Due in part to the low summer precipitation, the Mojave Desert lacks native grasslands, in contrast to the Chihuahuan Desert, where there are extensive grasslands dominated by C_4 species. In addition, two floristic seasons are often present in warm temperate deserts. In the deserts of the American Southwest, winter rains and mild temperatures support a rich flora of C_3 annuals that produce spectacular blooms during wet years. By contrast, summer rains produce a rich flora of C_4 annuals (Table VI). Because C_4 annuals typically produce inconspicuous flowers (Guo and Brown, 1997), the blooms of summer annuals fail to attract the attention that the cool-season blooms receive.

Moderate salinity tends to favor C_4 species, apparently because the inherently greater WUE reduces the salt load that the plants must deal with in the transpiration stream. In temperate to tropical latitudes, marshes and soils of moderate salinity are commonly dominated by C_4 species. For example, C_4 grasses of *Spartina* and *Distichilis* dominate salt marshes of coastal North America and Europe, and salinized rangelands of central Asia and western North America are often dominated by C_4 species in the Chenopodiaceae (see Table V). Moderate salinity also appears to allow C_4 species to dominate landscapes in climate zones that may otherwise be too cold. *Spartina* marshes predominate in eastern Canada and northern Europe at latitudes where inland there is no C_4 dominance. In turn, in Mediterranean climate zones, C_4 salt grasses (*Distichilis* spp.) can dominate coastal bluffs and cliff ledges because higher salinity allows them to compete with the local C_3 flora during the cooler months of the year.

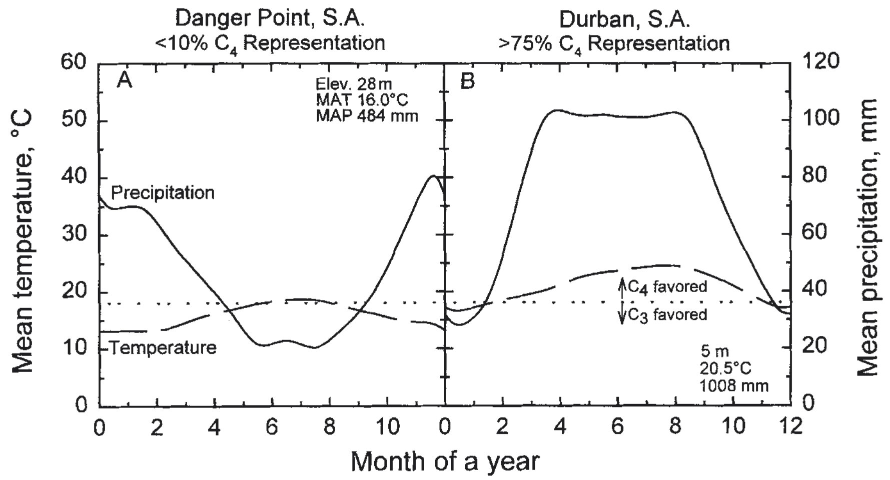


FIGURE 10 Walter-type climate diagrams for Danger-Point (near Capetown) and Durban, South Africa. Monthly average precipitation is indicated by solid lines, and monthly temperature means by dashed lines. The dotted line indicates the crossover temperature above which temperature favors C_4 grass vegetation over C_3. MAT, mean annual temperature; MAP, mean annual precipitation. (From Sage, R. F., and Pearcy, R. W. (2000). The physiological ecology of C_4 photosynthesis. In: *Photosynthesis: Physiology and Metabolism* (R. C. Leegood, T. D. Sharkey, and S. von Caemmerer, Eds.), pp. 497–532. Kluwer Academic, Dordrecht, Netherlands, with permission.)

IV. TAXONOMIC DIVERSITY

A. Description

There are 7000 to 8000 C_4 plant species. In proportional terms, 3 to 4% of the world's terrestrial flora is C_4. All C_4 species are angiosperms, with no known C_4 ferns, conifers, lower vascular plants, or algae. Of the C_4 angiosperms, all are in advanced rather than primitive orders. Recent angiosperm phylogenies place the presence of C_4 photosynthesis in seven taxonomic orders, with the Zygophyllaceae remaining unclassified to order because of uncertain affinity (Fig. 11).

The distribution of C_4 photosynthesis within the angiosperm orders is dispersed, with no apparent pattern. Two monocot and five dicot orders include C_4 families. Within the dicots, the Eurosids I and II lines and Euasterids I and II lines each include an order containing C_4 families. With the exception of the Caryophyllales (with eight families), each order has only one or two families containing C_4 species. In turn, many of these families include only a few C_4 genera. The most prolific C_4 families are the Poaceae (grasses) with almost 400 genera and 4600 species and the Cyperaceae (sedges) with 21 genera and about 1400 species (Table VII). The major C_4 dicot families are the Chenopodiaceae, Amaranthaceae, and Asteraceae, which together contain two-thirds of the C_4 dicot genera and species. With the exception of the Chenopodiaceae and Amaranthaceae, C_4 representation is generally low in the dicot families in which it occurs. Nearly 40% of the chenopods and 25% of the amaranth species are C_4, in contrast to 10 dicot families in which the C_4 syndrome is present in less than 5% of the species (Sage *et al.*, 1999a).

B. Polyphyletic Origins

The widespread occurrence of C_4 photosynthesis in diverse angiosperm lineages was an initial observation indicating that C_4 photosynthesis is polyphyletic and of recent origin. Further support for this interpretation came from comparative biochemical and anatomical studies that describe the variation in decarboxylation steps and leaf anatomy between different C_4 lineages. From this work, over 30 independent origins of C_4 photosynthesis have been identified (Kellogg, 1999).

Of the three decarboxylation pathways, the NADP-ME and NAD-ME pathways are evolutionarily unrelated, while the PCK pathway is derived from the NAD-ME pathway (Kellogg, 1999). Taxa with NAD-ME or PCK biochemistry are thus the result of separate evolutionary origins than NADP-ME taxa. On the basis of this understanding, it can be concluded that C_4 photosynthesis arose on multiple occasions within both families and individual genera. Of the 18 families with C_4

TABLE VI

Photosynthetic Pathway and Mean Density over a 10-Year Period for the Twelve Most Abundant Species of Summer and Winter Annuals on the Cave Creek Bajada, near Portal, Arizona (Chihuahan Desert)

Species	Photosynthetic pathway	Density (plants per 0.25 m^{-2})
Winter annuals		
Haplopappus gracilis	C_3	36.11
Eriastrum diffusum	C_3	30.65
Eriogonum abertianum	C_3	30.25
Descurania pinnata	C_3	4.07
Erodium cicutarium	C_3	3.77
Cryptantha micrantha	C_3	2.55
Gilia sinuata	C_3	2.01
Baileya multiradiata	C_3	1.56
Malacothrix fendleri	C_3	1.53
Spermolepis echinata	C_3	1.41
Plantago purshii	C_3	1.38
Erigeron divergens	C_3	1.34
Summer annuals		
Bouteloua aristidoides	C_4	6.92
Mollugo verticillata	C_4	3.82
Eriogonum abertianum	C_3	3.60
Mollugo cerviana	C_4	2.46
Pectis papposa	C_4	2.27
Aristida adscensionis	C_4	2.20
Boerhavia spicata	C_4	1.34
Chenopodium fremontii	C_3	0.87
Euphorbia fendleri	C_4	0.84
Baileya multiradiata	C_3	0.72
Tidestromia lanuginosa	C_4	0.65
Panicum arizonicum	C_4	0.44

[a] Developed from Guo and Brown (1996).

photosynthesis, 5 (Amaranthaceae, Chenopodiaceae, Cyperaceae, Poaceae, and Portulacaceae) have genera that include NADP-ME and NAD-ME species. In grasses, *Panicum* has all three biochemical subtypes; among dicots, two genera (*Salsola* and *Portulaca*) have the NAD-ME and NADP-ME subtypes, as well as containing C_3 members.

The extent to which C_4 photosynthesis is polyphyletic has become apparent in recent phylogenetic surveys. In the Cyperaceae, for example, four distinct C_4 lineages are present, each of which has a characteristic anatomy (Fig. 12). Fourteen C_4 sedge genera fall into the subfamily Cypereae, including the large genus *Cyperus*, which is about 60% C_4. Five genera are in the subfamily Abildgaardieae and two genera are in the Rhynchosporeae. The C_4 plants in each of these subfamilies are NADP-ME, but show distinct anatomical characteristics indicating independent C_4 origin (Soros and Bruhl, 2000). The fourth C_4 line in the Cyperaceae is represented by *Eleocharis* in the otherwise C_3 sub-family Scirpeae. *Eleocharis* is NAD-ME, and it also has a unique C_4 anatomy.

Current models indicate at least three independent origins for NADP-ME C_4 grasses, in the subfamilies Aristideae, Eriachne, and Panicoideae (Kellogg, 1999). Recent reports that *Orcuttia* in the Chloridoideae is NADP-ME indicate a fourth independent origin in the Poaceae (Keeley, 1998). The NAD-ME subtype originated independently in the Panicoideae and Chloridoideae, and may have multiple independent origins within these subfamilies, because in both there is considerable anatomical and cellular variation.

For reasons that are not known, the vast majority of C_3 families lack C_4 photosynthesis, even though many of them occur in environments where the C_4 pathway confers substantial benefits. Notably, most of the largest families of flowering plants lack C_4 photosynthesis. The orchids (Orchidaceae), rushes (Juncaceae), legumes (Fabaceae), and Rosaceae have not produced any C_4 taxa, despite sharing habitat and having close taxonomic affinity with some successful C_4 lines. For example, rushes and sedges are sister families, and herbaceous legumes are one of the few C_3 groups that routinely occur with C_4 grasses in low-latitude grasslands and deserts. Also for reasons unknown, few families contain both CAM and C_4 lines. Only five are known to have both CAM and C_4 genera, and only one of these, the Portulacaceae, has both CAM and C_4 species within the same genera (Table VIII). Only one species, *Portulaca oleracea*, is known to be both CAM and C_4, which is a sharp contrast to many CAM species that operate in the C_3 mode when water is abundant.

An important question is why C_4 photosynthesis is concentrated in a relatively small number of taxa. Are there characteristics that predisposed these groups to develop C_4 photosynthesis? Ehleringer *et al.* (1997) suggest that grasses and other species with close vein spacing are more likely to develop C_4 photosynthesis than species (dicots) with wider vein spacing. They believe that plants with close vein spacing may have a greater probability of establishing the C_4 pattern of metabolite transport between mesophyll and bundle sheath cells. The habitat of the C_3 progenitors to C_4 taxa is also an important consideration. C_3 grasses are common in arid landscapes of high light, and C_3 chenopods are common in saline habitats over much of the earth. Adaptations contributing to success in open, high-light, and high-stress habitats may have predisposed species of these families to evolve traits

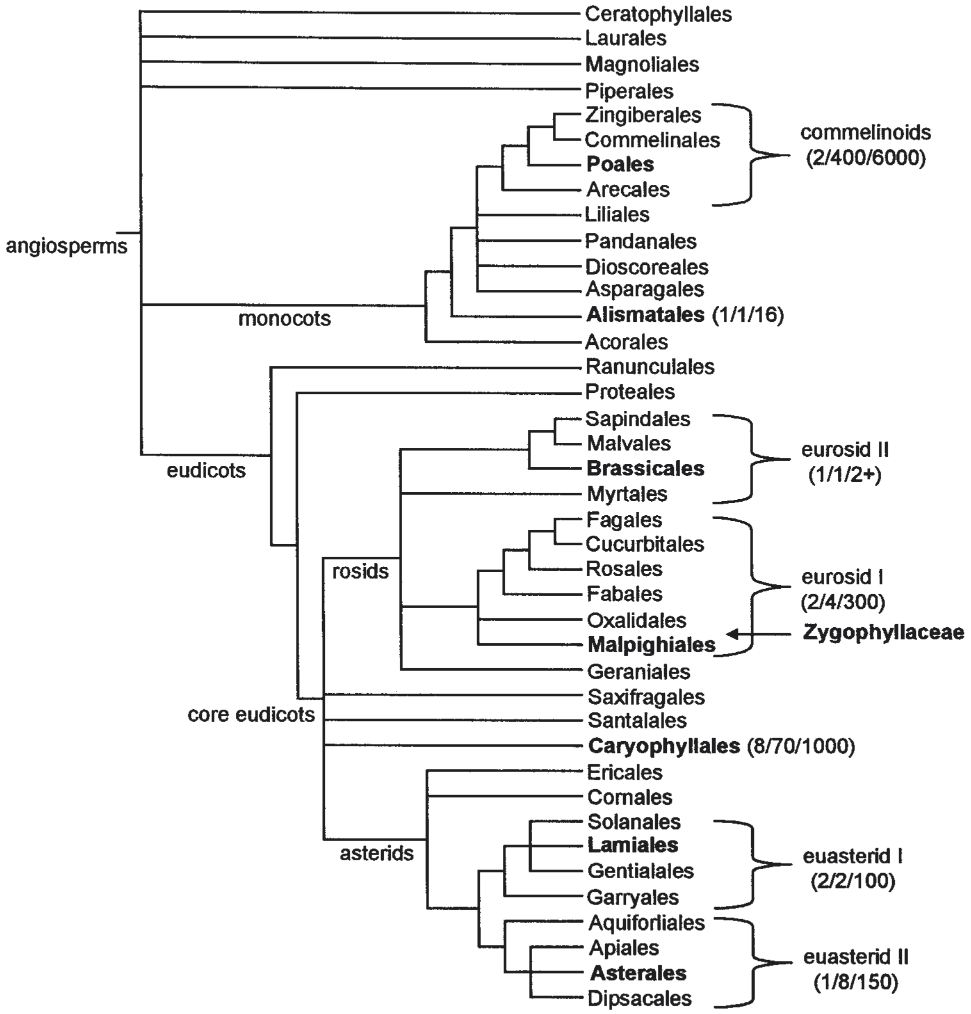


FIGURE 11 Phylogenetic relationships within the flowering plants, with C_4 orders and families showed in bold. Numerals in parentheses beside a group name indicate the number of C_4 families/genera and estimated number of species within the group. (Reprinted from the Angiosperm Phylogeny Group, 1998, with permission.)

that facilitated subsequent development of a CO_2 concentration system. In addition, the reduction in biotic competition often found in severe environments may have enabled many intermediate forms to evolve, with some eventually developing the complete C_4 pathway.

V. THE RISE OF C_4 PHOTOSYNTHESIS

A. When Did C_4 Plants Evolve?

Three lines of evidence are used to evaluate when C_4 photosynthesis first appeared. These are: (1) fossilized grasses with Kranz anatomy and other characteristics found in modern C_4 taxa; (2) carbon isotopic ratios from soils, plant remains, or herbivore fossils that demonstrate C_4 presence on the landscape; and (3) phylogenetic estimates of divergence times for current C_4 taxa. The first two lines of evidence demonstrate presence of C_4 plants on a landscape but do not demonstrate actual time of origins. Phylogenetic analyses potentially indicate exact times of origin, assuming the biomolecular clocks used in divergence estimates are accurate.

1. Fossil Evidence

The oldest known C_4 fossils come from the Ricardo formation of southern California, dated from the middle

TABLE VII

The Occurrence of C_4 Photosynthesis in Higher Plant Orders and Families[a]

		Genera statistics			Species statistics			
		# C_4	Total	% C_4	# C_4	Total	% C_4	% of all C_4
Dicotyledoneae (subclass)								
Asterales (order)	Asteraceae	8	1500	1%	150	13,000	1.3%	2.0%
Brassicales	Capparidaceae	1	34	3%	2+	925	<1%	<0.2%
Caryophyllales	Aizoaceae	5	126	4%	~30	2500	1.2%	0.4%
	Amaranthaceae	13	74	18%	~250	1000	25%	3%
	Caryophyllaceae	1	88	1%	50	2200	2%	0.6%
	Chenopodiaceae	45	105	43%	~550	1400	39%	7%
	Molluginaceae	2	14	14%	4+	120	3%	0.2%
	Nyctaginaceae	3	33	9%	5+	400	1%	0.3%
	Portulacaceae	2	20	10%	70	450	16%	0.9%
Euphorbiales	Euphorbiaceae	1	300	<1%	250	5000	5%	3%
Linnales	Zygophyllaceae	3	30	10%	~50	240	21%	0.7%
Polygonales	Polygonaceae	1	45	2%	80	1100	7%	1%
Solanales	Boraginaceae	1	250	<1%	6+	2000	<1%	<0.2%
Scrophulariales	Acanthaceae	1	250	<1%	80	2400	3%	1%
	Scrophulariaceae	1	280	<1%	14	3000	<1%	<0.2%
Total dicot	**15**	**90**	**3150**		**~1600**	**35,735**		**21**
Monocotyledoneae								
Alismatales	Hydrocharitaceae	1	16	6%	1+	100	1%	<0.1%
Juncales	Cyperaceae	28	131	21%	1330	5000	27%	18%
Poales	Poaceae	372	~800	47%	4600	10,000	46%	61%
Total monocot	**3**	**401**	**947**		**~6000**	**15,000**		**79%**
Total C_4	**18**	**491**			**~7600**			**100%**

[a] Genera statistics are the number of genera having C_4 species, total number of genera in the family, and percentage of genera in the family with C_4 members. Species statistics refer to total number of C_4 species estimated for the family, total species in the family, and percentage of those species that are C_4. The percentage of all C_4 refers to the number of C_4 species in a taxonomic group divided by 7600, the estimated total number of C_4 species. Estimates are from Sage *et al.* (1999a). A "+" after the species number indicates an underestimate.

Miocene at approximately 12.5 million years ago (mya) (Cerling, 1999). These fossils show distinct C_4 Kranz-type leaf anatomy that is characterized by enlarged bundle sheath cells and reduced numbers of mesophyll cells. In addition, carbon extracted from the samples has isotopic ratios indicative of C_4 species. The C_4 grasses at this location appear panicoid, and are apparently not common, as only a small number of samples have been found despite repeated searching (Tidewell and Nambudiri, 1989). The only other definitive C_4 fossils known from pre-Pliocene deposits are from the Ogalla formation of northwestern Kansas (U.S.A.), dated at 5 to 7 mya (Thomasson *et al.*, 1986). These fossils show distinct C_4 leaf anatomy, and in other features closely resemble extant species in the Chloridoideae, a nearly pure C_4 grass subfamily.

Probable C_4 plant remains have been found in the Fort Ternan formation of Kenya, dated at 14.5 mya (Cerling, 1999). Because internal tissues have been lost in these remains, only morphological and cuticular features are useful. These indicate that the plants were Chloridoideae and Panicoideae grasses. The Chloridoideae specimens match modern C_4 genera, and one has been identified as *Distichilis*, a widespread C_4 genus. Of the Panicoideae fossils, one matches the C_4 genus *Stereochleana*, which is now common in East Africa (Dugas and Retallack, 1993).

2. Carbon Isotope Discrimination

In contrast to the sparse record for early C_4 plant fossils, there is abundant isotopic evidence documenting a widespread radiation of C_4 plants at the end of the Miocene period 5 to 8 million years ago. C_4 photosynthesis discriminates less against heavy isotopes of carbon than does C_3 photosynthesis, such that C_4 tissues have 8 to 15 more ^{13}C atoms per hundred thousand ^{12}C atoms than do C_3 plants (see Table II). Using mass spectrometry, these differences can be readily detected in plant remains, or any material that originated from photosynthetic carbon.

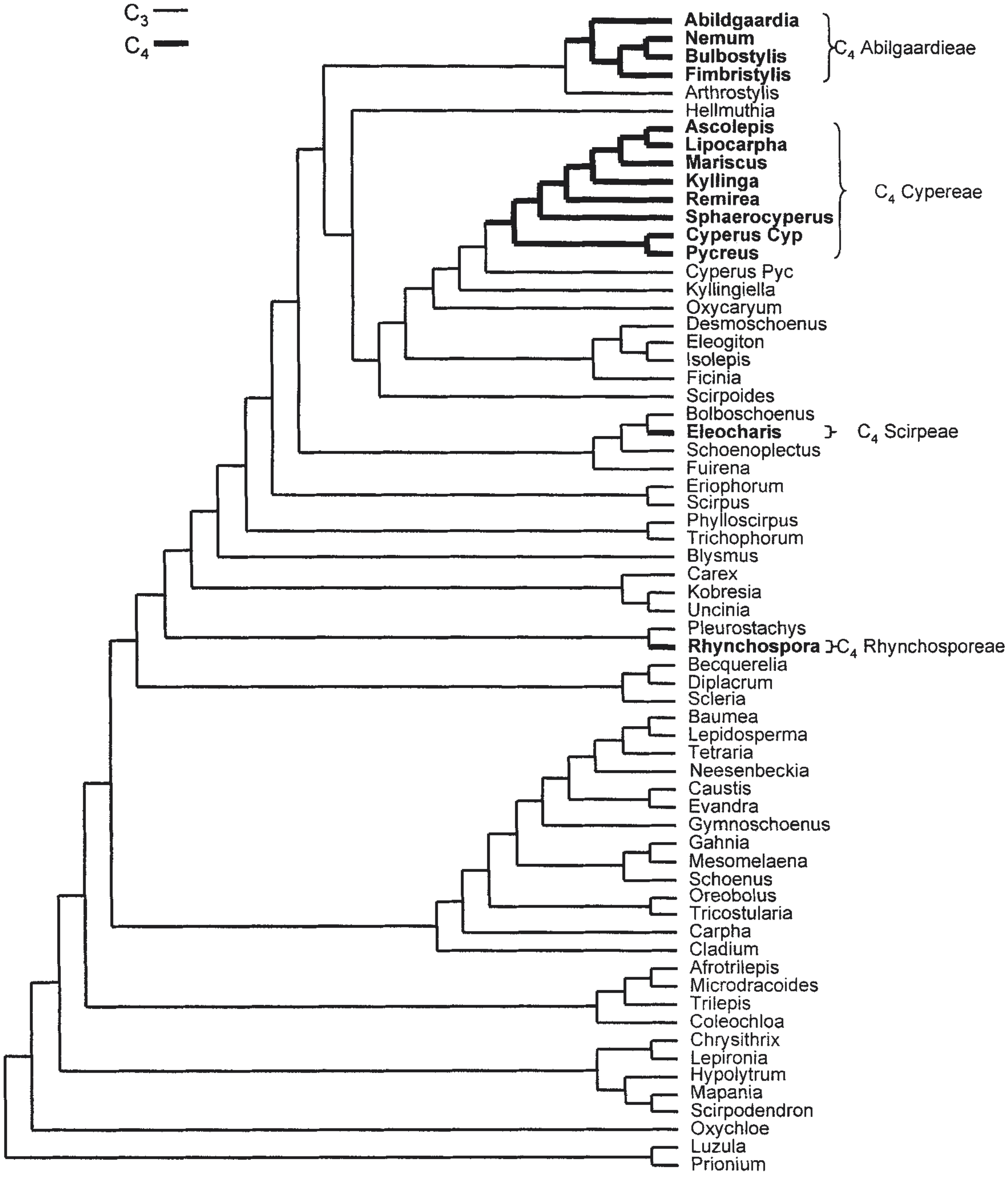


FIGURE 12 The distribution of C_4 photosynthesis within the Cyperaceae. Bold lines indicate C_4 lineages. Names beside brackets indicate the subfamilies containing C_4 genera. The cladogram is derived from both molecular and nonmolecular data of a subset of genera within the family. Not all C_4 genera are shown. (From Soros and Bruhl, 2000, with permission.)

Thus, fossilized organic matter, soil carbonates, and animal remains can be assessed to determine whether C_3 or C_4 species were predominant on the landscape at the time of fossilization.

Both soil and herbivore remains from around the world show a clear shift in isotopic ratios of 5 to 10 per mil between 5 and 10 mya (Fig. 13). In East Africa and Pakistan, isotopic shifts are dramatic, indicating near-complete replacement of C_3 with C_4 vegetation in as little as 2 million years at many sites. At higher

TABLE VIII

Plant Families with Both C_4 and CAM Species[a]

Family	Total genera	C_4 genera	CAM genera	C_4 and CAM genera
Aizoaceae	126	5	47	0
Asteraceae	1500	8	4	0
Euphorbiaceae	300	1[b]	4	0[b]
Hydrocharitaceae	16	1	1	0
Portulacaceae	20	2	6	2[c]

[a] Compiled from Sage *et al.* (1999a) and Smith and Winter (1996).

[b] C_4 members of the Euphorbiaceae are grouped here into the genus *Chaemasyce*. Some treatments place these species into *Euphorbia*, a genus with CAM and C_3 species.

[c] The C_4 and CAM genera are *Anacampseros* and *Portulaca*. Only *Portulaca* is known to have species that switch between C_4 and CAM modes.

latitude, substantial variation is present in isotopic data after about 7 million years, indicating partial replacement of C_3 with C_4 species.

3. Phylogenetic Approaches

Fossil and isotopic analyses can detect when C_4 plants are well represented in regions where preservation can occur, but they may not detect actual times of origin if the species remain rare or occur in places such as arid zones where preservation is unlikely. To overcome this problem, molecular phylogenies from extant genera can be used to identify divergent times between C_3 and C_4 taxa. This approach relies on predictable rates of divergence in the DNA, RNA, and protein sequences of related C_3 and C_4 material. Using this approach, C_4 photosynthesis is estimated to have emerged over 25 mya (Kellogg, 1999). Maize (*Zea mays*) appears to have diverged from *Pennisetum* by 25 mya, and from sorghum (*Sorghum halopense*) by 18 mya. Because all three species are NADP-ME grasses sharing a common C_4 ancestor, their divergence is believed to have occurred after the origin of the C_4 pathway.

In contrast to grasses, there is little evidence for the origins of C_4 dicots. C_4 dicot fossils are unknown, and phylogenetic divergence estimates have yet to be presented. C_4 dicots will be a valuable source of enquiry into C_4 origins, given the high number of distinct, and apparently recent, origins. Based on the low number of C_4 dicot taxa and multiple appearances, Ehleringer *et al.* (1997) speculate that C_4 dicots evolved more recently than C_4 monocots, with most origins possibly occurring within the past 3 million years.

All current families containing C_4 photosynthesis arose no earlier than the late Cretaceous (65 to 75 mya). Grasses and sedges appear at the end of the Cretaceous period, and become common during the Eocene epoch (58 to 36 mya). Most dicot families that now contain C_4 species appear in the fossil record between 50 to 20 mya, a time when most orders and families of angiosperms arose. The Chenopodiaceae/Amaranthaceae complex appears to be the oldest of the large C_4 dicot families, with the earliest pollen dating to the late Cretaceous about 68 mya. Early Chenopodiaceae fossils are associated with marine/terrestrial boundary zones, indicating saline habitats. Where fossil material is present, the first members of these families appear to be C_3 species, with C_4 species appearing in more advanced lineages (Kellogg, 1999).

The origin of the plant families that now contain C_4 species is associated with a global cooling phase that occurred during the middle Tertiary period between 50 and 20 mya. In low to middle latitudes, this climatic shift produced semiarid landscapes that by the middle Miocene (15 mya) were supporting grasslands and savannas that possibly contained isolated swards of C_4 grass. Similarly, sedges, chenopods, and members of other families that eventually developed C_4 species become more common after 35 mya (Collinson *et al.*, 1993).

B. Driving Forces for the Evolution of C_4 Species

On the basis of the taxonomic and phylogenetic evidence, a likely scenario is that C_4 photosynthesis originated between the middle Oligocene and early Miocene (between 35 and 18 mya) when global aridification became pronounced, while the subsequent expansion to dominance occurred during the late Miocene about 8 mya. The identification of these times is important, because it allows us to address hypotheses concerning the environmental factors that promoted the rise of C_4 photosynthesis.

The modern association of C_4 plants with hot, dry conditions has led to suggestions that C_4 photosynthesis evolved in response to climate warming and aridification. Though the postulated origin of C_4 photosynthesis in the middle to late Tertiary is consistent with paleoclimatic shifts to drier conditions, the global climate then was in general cooler than in previous periods. During the Cretaceous period (145 to 65 mya), for example, the earth was considerably warmer than in the middle to late Tertiary, but as far as we know there were no C_4 plants. Aridification alone is an inadequate explanation for the evolution of C_4 plants, because significant arid regions also existed long before C_4 species arose.

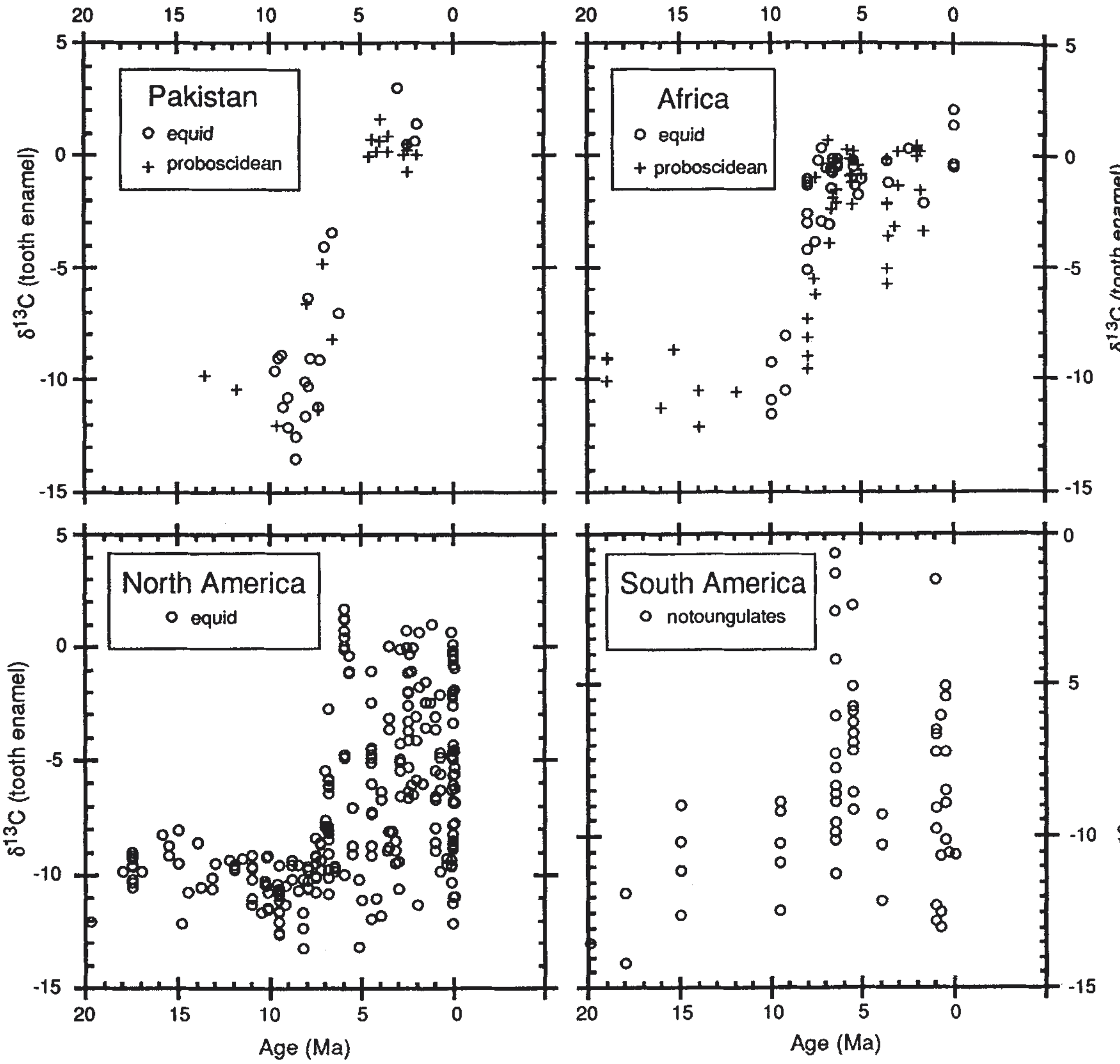


FIGURE 13 Carbon isotope ratios from fossil mammals over the past 20 million years. The shift in values from more negative to less negative values between 10 and 5 million years ago indicates the spread of C$_4$ vegetation in these localities at this time. (Reprinted from Cerling, 1999, with permission.)

Moreover, in hot climates, modern distributions indicate that aridity is not a prerequisite for C$_4$ photosynthesis, as long as disturbances are present to check the woody vegetation. During the Cretaceous period, dinosaurs would have provided substantial disturbance, and geological cores demonstrate that there was an abundance of fire.

One important parameter that changed during the period when C$_4$ plants evolved is the atmospheric level of CO$_2$. During the Cretaceous, atmospheric CO$_2$ levels were 5 to 10 times higher than at present, and as a result the potential was low for photorespiration in C$_3$ plants (Fig. 14). Since then, CO$_2$ levels have declined, and with this the potential for photorespiration has increased. The origin of C$_4$ species corresponds to the period when atmospheric CO$_2$ levels are predicted to have fallen to the point where photorespiratory potentials at 30°C rise above 20% of photosynthesis (see Fig. 14). This reduction in CO$_2$ is proposed to have been the driving force for the origin of C$_4$ species.

A CO$_2$ control over the timing of C$_4$ origins is attractive because it unifies the various lines of evidence concerning C$_4$ evolution. Physiological studies demonstrate no major advantage of C$_4$ photosynthesis at elevated CO$_2$; instead, the extra energy cost of C$_4$ photosynthesis should be detrimental except at light saturation. Both growth and photosynthesis measurements predict a shift from C$_3$ to C$_4$ dominance at CO$_2$ levels modeled to have existed during the late Miocene when C$_4$ species expanded across tropical landscapes. Aridity

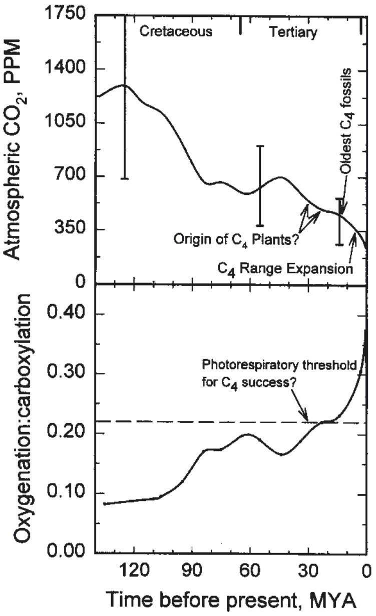


FIGURE 14 (A) Atmospheric CO_2 content modeled over the past 140 million years, with important times for the advent of C_4 species indicated. Error bars indicate the range of uncertainty in the CO_2 estimates. Developed from data in Berner (1994). (B) The theoretical oxygenation to carboxylation ratio corresponding to the CO_2 estimates in panel A and assuming 20% O_2 in the atmosphere and 30°C. (From Sage, 1999.)

likely played a contributing role because stomatal closure to conserve water also reduces the supply of CO_2 for photosynthesis.

Where did the first C_4 species evolve? Based on physiological models and current biogeographic patterns, it is likely that C_4 photosynthesis first appeared in hot, arid, and saline situations. Such environments now have the greatest diversity of C_4 taxa, with a wide range of dicot and monocot species, including most woody C_4 species. As CO_2 levels declined during the Tertiary period, these settings would have been the first to favor CO_2 concentration systems, presumably because the combination of heat, drought, and salinity would have provided the greatest inhibition of nearby C_3 competitors.

VI. CLIMATE CHANGE AND THE FUTURE OF C_4 PHOTOSYNTHESIS

If C_4 species arose because of past global changes that favored photosynthetic CO_2-concentrating mechanisms, then it follows that anthropogenic increases in CO_2 would reduce the advantage that C_4 species had in the low-CO_2 atmospheres of recent geological time. Cooler, more temperate environments would be affected first, with C_4 species in warm, arid habitats being the least affected. It is unlikely, however, that widespread extinction of C_4 species will result from atmospheric CO_2 rise.

First, global warming accompanying CO_2 rise will offset advantages that the CO_2 increase might confer upon C_3 plants. C_4 species are generally better adapted to warmer climates than potential C_3 competitors, and thus will be in a better position to exploit additional warmth. Second, feedbacks are present in plants and ecosystems that constrain responsiveness to high CO_2. Central among these is nutrient availability. If nutrients are deficient, the ability of C_3 species to exploit CO_2 enrichment is substantially reduced, if not lost altogether. Most natural ecosystems are nutrient deficient, particularly those dominated by C_4 species.

Third, a common limitation in ecosystems where C_4 species occur is water. In water-limited situations, productivity of C_3 and C_4 species is enhanced by rising CO_2 because stomata of both close as atmospheric CO_2 rises, leading to water savings and longer growing seasons. The severity of drought is often greater during warm summer months when evaporative demand is greatest and C_4 species are more active. Thus, positive responses to high CO_2 are likely to be greater in summer-active C_4 species rather than the spring-active C_3 species from the same community. Warmer temperatures and more frequent aridity will also promote fires that strongly favor C_4 species because of negative effects on the woody C_3 vegetation.

Fourth, changes in seasonality will be critical to future C_3/C_4 dynamics. Where winters become milder, yet remain cool, C_3 species could be favored. Such a pattern has been observed in central Canada, where a recent rise in early spring temperatures has been associated with increased C_3 grass cover at the expense of C_4 species. Increased summer precipitation in arid regions would favor C_4 species by providing necessary moisture for growth during the high-temperature period favorable to C_4 species. Thus, Mediterranean zones that develop a summer monsoon could see a dramatic rise in the cover of C_4 species.

Because past climate change has been associated with CO_2 rise, paleoecological studies indicate how future C_3/C_4 dynamics might develop. Between 20,000 and 10,000 years ago, regional climates warmed 5° to 10°C and CO_2 levels rose 30%, thereby providing a similar situation to that predicted to occur in coming centuries. Theoretically, the CO_2 rise should have had a dramatic effect on C_4 relative to C_3 distributions, because the low-CO_2 level of 20,000 years ago (180 ppm) would have been far more limiting for the performance of C_3 relative to C_4 plants. Paleoecological studies indicate that in tropical climates, there was a wider distribution of C_4 species 20,000 years ago relative to 10,000 years ago and today. In central Africa and South America, isotopic and pollen data indicate that grasslands were more widespread 20,000 years ago than at present, which is consistent with a low-CO_2 effect favoring C_4 plants. However, these regions were drier than at present, which could also explain the pattern.

In temperate zones, little CO_2 effect on C_3 versus C_4 plant distribution is apparent during this time. In southwestern North America, for example, the distribution of C_3- and C_4-dominated grass and shrub vegetation appears to remain stable over the past 20,000 years, possibly because the seasonality of precipitation has also changed little over this time (Connin *et al.*, 1999). In contrast to the American Southwest, substantial expansion of C_4 grasses occurred at the end of the last ice age on the plains grasslands of central North America. Here, C_4-dominated grasslands expanded northward between 12,000 and 8000 years ago, a time of little CO_2 change, but considerable climate warming.

VII. C_4 VEGETATION AND GLOBAL BIODIVERSITY

A. C_4 Grasses and Tropical Deforestation

Through the use of fire, humans have been major agents of vegetation change for over 2 million years. The most significant impact of burning has come in the past few thousand years, when humans increasingly used fire to maintain and expand agricultural lands. As a result, many of the grasslands and savannas of the world are derived from human-caused fires. In recent decades, expansion of C_4-dominated grasslands has become pervasive in tropical regions, and in the process has become one of the more important threats to tropical biodiversity. This expansion has largely been driven by human population and economic growth, but the introduction of alien C_4 grasses into heavily impacted systems has promoted synergistic effects that often convert high-diversity forests to low-diversity grasslands.

Historically, agricultural practices in tropical regions consisted of shifting cultivation, in which forest plots would be slashed and burned, and the resulting open ground farmed for a few years and then abandoned. Plots were generally small enough and surrounding forest large enough to restrict the access of weedy grasses while providing large numbers of propagules from the nearby woodland. This ensured that the plots would readily undergo succession back to forest. In recent decades, however, higher human populations, technological improvements such as the use of chain saws, and high demand for timber and beef exports have promoted the widespread exploitation of tropical landscapes. As a result, the scale of land clearing has become far more extensive, with open areas favorable to C_4 grasses being produced at an increasing rate, while forests are becoming increasingly fragmented and influenced by the drier microclimates of the grassland areas. Coupled with these trends in land exploitation has been the introduction of cultivated or weedy C_4 grasses that are well adapted to human activities. In particular, aggressive growth of the weedy invaders following human disturbance has caused shifts in fire cycles, soil biogeochemistry, and food webs so that succession back to forest is often arrested and low-diversity grasslands become permanent fixtures on the landscape.

Numerous activities cause the deforestation of tropical forests, but three stand out as the major causes for forest decline. In each case, C_4 grasses play a key role.

1. Direct Clearing of Forest for Pasture

Forest clearing for pasture is particularly pervasive in Amazonia, where demand for low-cost beef is driving the expansion of the cattle industry into forested regions. Typically, primary forest with plant species numbering in the hundreds to thousands per hectare are converted to managed pastures with less than 10 plant species per hectare, most of which are economically useful C_4 grasses and their weedy associates.

2. Runaway Slash and Burn Agriculture

As human populations in forested areas have grown, slashed and burned plots have become larger and rotation times shorter. Invasive C_4 weeds become severe problems as these farming plots age, contributing to their abandonment after a few years. These weedy grasses are tough competitors, not only because they have high production rates, but also because their leaves decompose slowly due in part to having higher carbon-to-nitrogen ratios than C_3 leaves. Slow decomposition reduces soil nitrogen availability, thereby favoring the C_4 species that have higher nitrogen-use efficiency.

Many of the weedy C_4 invaders are pyrophytic; not only do they survive fire well, but they rapidly produce a dense canopy that readily dries to produce a highly flammable fuel. During dry episodes, the grass swards rapidly burn, producing fires that are lethal not only for forest seedlings establishing in the swards, but also for adjoining patches of forest. Once established, exotic grass swards will typically burn every 2 to 5 years. With each burn, more forest is destroyed, and fewer remnant forest species are present in the grassland. After a few cycles of burning, the forest seed bank is destroyed, and forest patches are too isolated to contribute to the seed rain over most of the grassland. At this point, the grassland may become a permanent fixture on the landscape.

Compared to the original diversity of hundreds to thousands of plant species per hectare, weedy grasslands are dominated by a few species of C_4 grass, although they may include up to a few dozen species of minor herbaceous plants. In the extreme, weedy grasses can form dense stands that exclude all other plant species. In Southeast Asia and Indonesia, the C_4 grass *Imperata cylindrica* has become a widespread pest because of its ability to rapidly colonize wet soils cleared for farming or grazing, after which it forms highly flammable swards that kill all other species through crowding or burning. In Central and South America, the C_4 grasses *Melinus Minutiflora*, *Panicum maximum*, and various *Pennisetum* species are causing loss of regional diversity for similar reasons.

3. Deforestation following Selective Logging of Tropical Forests

The demand for wood products and fuel has led to extensive logging operations throughout the wet tropics. Much of this logging is selective in nature, causing the forest to become disturbed but not destroyed. Although a selectively logged forest can succeed back to closed-canopy stands within a century, any logging increases the probability that they will burn owing to the increased level of downed woody debris (Fig. 15). Once burned, the level of woody debris increases as vegetation killed by the fire falls to the ground. This promotes a second fire within a few years, particularly if grasslands are nearby. Each fire tends to be more intense than the previous because of the increased level of woody debris, and each fire increases the probability of a subsequent burn. Critical to this cycle is the establishment of pyrophytic C_4 grasses in forest margins and gaps caused by repeated burning. These further accelerate fire cycles because the microclimate over the grass swards is warmer and less humid, thereby drying adjoining forests, and the grasses produce an abundance of fuel that carries fire into forest fragments.

The C_4 grass species contributing to biodiversity impoverishment across the globe are relatively small in number, and are disproportionately African in origin (Table IX). Part of the reason why African species are so successful is because they have had a long time to adapt to human activities. In contrast, grasses in the Americas and on oceanic islands have been exposed to humans for less than 14,000 years, and to intensive grazing regimes for only a few hundred years.

In addition to adaptation to humanity, unique features arising from C_4 photosynthesis help explain the expansion of these grasses at the expense of native species. C_4 photosynthesis promotes the high levels of primary production that make C_4 crops and forages so valuable to humans in warm tropical climates. This high productive potential enables C_4 grasses to rapidly establish in tropical locations relative to their C_3 competitors, such that they become the preferred pasture species, as well as the most aggressive weeds. The high water-use efficiency of C_4 grasses enhances tolerance of high evaporative demand and periodic drought that develops rapidly in tropical settings, typically after only a few days of no rain. The high nitrogen-use efficiency of C_4 grasses provides for an ability to produce greater amounts of biomass on infertile soils. Soil infertility often follows forest clearing because of leaching during heavy rainfall, nitrogen volatilization during fire, and nutrient exhaustion during cultivation. Importantly, the combination of these features allows C_4 grasses to rapidly produce large quantities of biomass that will soon provide fuel to promote intense, destructive fires.

B. Impoverishment of C_4-Dominated Landscapes

While certain C_4 species that are well adapted to human activity have greatly expanded their range in recent decades, most natural C_4-dominated ecosystems have been severely impacted by human activity. Many of the temperate C_4 grasslands of the world have already been destroyed, largely by human exploitation for agricultural production, fire suppression, or conversion to managed timber plantations. In North America, about 90% of the once widespread tallgrass prairie biome has been converted to row crop agriculture, forest, or pasture. Where stands of bluestem and switchgrass (*Schizychyrium*, *Andropogon*, and *Panicum* spp.) once formed an ocean of C_4 grass from Illinois to Kansas in the central United States, maize and soybean fields now stretch to the horizon. In the Southeast, a wiregrass/long-leaf pine savanna once extended across as much as 60% of the coastal plain from Virginia to Texas. Now, about 93% of this ecosystem has been lost to timber plantations, crop fields, and pasture (Drew *et al.*, 1998). In the western steppes and deserts, overgrazing coupled with climate change has contributed to conversion of

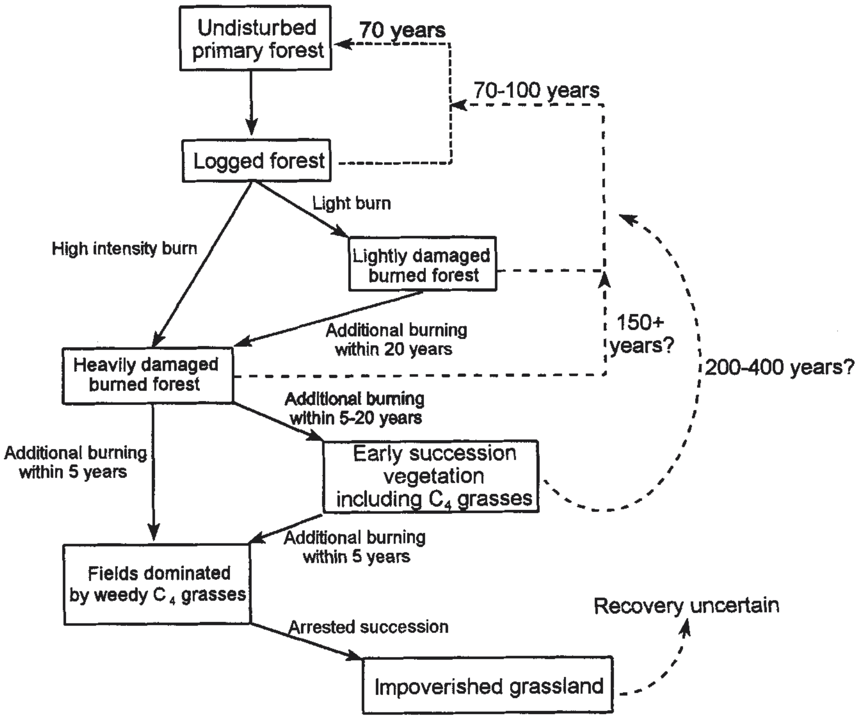


FIGURE 15 A conceptual model of forest conversion to grassland following logging. Dashed lines indicate recovery times if fire is excluded from the system. Arrested succession occurs by repeated burning, loss of soil fertility, and/or loss of the forest seed bank. (From Cochrane and Schulze, 1999, with modification to emphasize the role of C_4 grasses.) Copyright 1999 by the Association for Tropical Biology, P.O. Box 1897, Lawrence, KS 66044, USA. Reprinted by permission.

TABLE IX

Important Invasive C_4 Grasses of Natural Landscapes of the World[a]

Invaded region	Invaders	Region of origin
North America	*Microstegium vimineum*	Asia
	Eragrostis cuvula, Eragrostis lehmanniana, Cynodon dactylon, Digitaria spp., *Echinochloa* spp., *Pennisetum* spp., *Sorghum halopense*	Africa
Tropical America	*Brachiaria* spp., *Hyparrhenia rufa, Melinis minutiflora, Panicum maximum*	Africa
Southeast Asia	*Imperata cylindrica*	Africa
Australia	*Cenchrus ciliaris, Eragrostis curvula, Melinis minutiflora, Pennisetum polystachyon*	Africa
Oceania (Hawaii)	*Andropogon virginicus, Schizachyrium condensatum*	North America
	Cenchrus ciliarus, Digitaria decumbens, Hyparrhenia rufa, Imperata cylindrica, Melinus minutiflora, Pennisetum clandestinum, Pennisetum purpureum	Africa
Europe, New Zealand	*Spartina anglica* (= *S. townsendii*) in coastal salt marshes of Europe (this is a novel hybrid between *S. alterniflora* from the Americas and *S. maritima* from Europe)	

[a] Reprinted from Sage *et al.* (1999b), with permission.

much of the arid C_4 grasslands to C_3 shrublands. The desert grasslands of Texas and New Mexico have been particularly affected, with most of the original grassland now converted to mesquite (*Prosopis*) and *Acacia* scrub. Although half of the shortgrass prairie in the high plains of the central states has been converted to dryland or irrigated farms, large expanses remain intact, but often in a degraded condition due to overgrazing.

Worldwide, similar patterns hold. Most productive areas once dominated by C_4 grasslands have already been developed into cropland, pasture, or timber plantations, while less productive areas are routinely subjected to unsustainable levels of grazing. Recent estimates suggest that 72–74% of the rangelands of Africa, Asia, and the Americas are in a moderate to very severe degradation category; in Australia, 55% of the rangelands are degraded (McNeely *et al.*, 1995). Most of these rangelands once supported extensive C_4 grasslands. Land renewal efforts are under way in many of these areas, but the main focus is to establish a few productive species for pasture or wood production, rather than restoration of native C_4-dominated ecosystems. Although this process differs from that driving deforestation, the ultimate cause—unsustainable exploitation by humanity—and the end result—species impoverishment—remain the same.

VIII. SUMMARY

C_4 photosynthesis is a polyphyletic solution to challenges imposed by reductions in atmospheric CO_2 and increasing aridification that occurred over the past 50 million years. In C_3 plants, the low atmospheric CO_2 levels of recent geological time promoted substantial inhibition of photosynthesis by photorespiration in hot conditions, particularly if water was also limiting. C_4 plants overcame these limitations by localizing Rubisco in the bundle sheath compartment into which CO_2 is concentrated. As a consequence, C_4 plants are highly productive, aggressive competitors under conditions promoting photorespiration in C_3 plants. This improved performance has enabled native C_4 species to dominate hot, open landscapes of low latitudes, and warm temperate regions receiving summer rain. It has also enabled invasive C_4 grasses adapted to human disturbance to become severe weeds, with important consequences for tropical biodiversity. Where humans disrupt forest cover and allow grass establishment, they establish conditions where invasive C_4 grasses can accelerate fire cycles and alter soil chemistry. In extreme cases, diverse tropical forests can be replaced by grasslands dominated by a few aggressive C_4 species. Because natural C_4-dominated grasslands and savannas often occur on more productive landscapes, they too have been heavily exploited for agricultural purposes, resulting in habitat loss and impoverishment of the native C_4 flora of the planet.

See Also the Following Articles

CARBON CYCLE • DEFORESTATION • DESERT ECOSYSTEMS • MEDITERRANEAN-CLIMATE ECOSYSTEMS • NITROGEN AND NITROGEN CYCLE • PHOTOSYNTHESIS, MECHANISMS OF • SLASH AND BURN FARMING

Bibliography

Angiosperm Phylogeny Group. (1998). An ordinal classification for the families of flowering plants (K. Bremer M. W. Chase, and P. F. Stevens, eds.). *Ann. Missouri Bot. Garden* 85, 531–553.

Berner, R. A. (1994). 3Geocarb II: A revised model of atmospheric CO_2 over Phanerozoic time. *Amer. J. Sci.* 291, 339–376.

Brown, R. H. (1999). Agronomic implications of C_4 photosynthesis. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 473–507. Academic Press, San Diego.

Cerling, T. E. (1999). Paleorecords of C_4 plants and ecosystems. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson eds.), pp. 445–469. Academic Press, San Diego.

Cerling, T. E., Harris, J. M., MacFadden, B. J., Leacey, M. G., Quade, J., Eisenmann, V., and Ehleringer, J. R. (1997). Global vegetation change through the Miocene/Pliocene boundary. *Nature* 389, 153–158.

Cochrane, M. A., and Schulze, M. D. (1999). Fire as a recurrent event in tropical forests of the eastern Amazon: Effects on forest structure, biomass, and species composition. *Biotropica* 31, 2–16.

Collatz, G. J., Berry, J. A., and Clark, J. S. (1998). Effects of climate and atmospheric CO_2 partial pressure on the global distribution of C_4 grasses: Present, past and future. *Oecologia* 114, 441–454.

Collinson, M. E., Boulter, M. C., and Holmes, P. L. (1993). Magnoliophyta ("Angiospermae"). In *The Fossil Record* 2 (M. J. Benton, ed.), pp. 809–841. Chapman and Hall, London.

Connin, S. L., Betancourt, J., and Quade, J. (1999). Late Pleistocene C_4 plant dominance and summer rainfall in the southwestern United States from isotopic study of herbivore teeth. *Quat. Res.* 50, 179–193.

D'Antonio, C. M., and Vitousek, P. M. (1992). Biological invasions by exotic grasses, the grass/fire cycle, and global change. *Annu. Rev. Ecol. Syst.* 23, 63–87.

Dengler, N. G., and Nelson, T. (1999). Leaf structure and development in C_4 plants. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 133–172. Academic Press, San Diego.

Drew, M. B., Kirkman, L. K., and Gholson, A. K., Jr. (1998). The vascular flora of Ichauway, Baker County, Georgia: A remnant longleaf pine/wiregrass ecosystem. *Castanea* 63, 1–24.

Dugas, D. P., and Retellack, G. J. (1993). Middle Miocene fossil grasses from Fort Ternan, Kenya. *J. Paleontol.* 67, 113–128.

Ehleringer, J. R., and Monson, R. K. (1993). Evolutionary and ecological aspects of photosynthetic pathway variation. *Annu. Rev. Ecol. Sys.* 24, 411–439.

Ehleringer, J., and Pearcy, R. W. (1983). Variation in quantum yield for CO_2 uptake among C_3 and C_4 plants. *Plant Physiol.* 73, 555–559.

Ehleringer, J. R., Cerling, T. E., and Helliker, B. R. (1997). C_4 photosynthesis, atmospheric CO_2, and climate. *Oecologia* 112, 285–299.

Groombridge, B. (ed.). (1992). *Global Biodiversity: Status of the Earth's Living Resources.* Chapman and Hall, London.

Guo, Q., and Brown, J. H. (1997). Interactions between summer and winter annuals in the Chihuahuan desert. *Oecologia* 111, 123–128.

Hatch, M. D. (1987). C_4 photosynthesis: A unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* 895, 81–106.

Kanai, R., and Edwards, G. E. (1999). The biochemistry of C_4 photosynthesis. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 49–87. Academic Press, San Diego.

Keeley, J. E. (1998). C_4 photosynthesis modification in the evolutionary transition from land to water in aquatic grasses. *Oecologia* 116, 85–97.

Kellogg, E. A. (1999). Phylogenetic aspects of the evolution of C_4 photosynthesis. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 411–444. Academic Press, San Diego.

Larcher, W. (1995). *Physiological Plant Ecology*, 3rd ed. Springer, Berlin.

Long, S. P. (1999). Environmental responses. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 215–249. Academic Press, San Diego.

McNeely, J. A., Gadgil, M., Leveque, C., Padoch, C., and Redford, K. (1995). Human influences on biodiversity. In *Global Biodiversity Assessment* (V. H. Heywood and R. T. Watson, eds.), pp. 711–821. Cambridge University Press, Cambridge, United Kingdom.

Monson, R. K. (1999). The origins of C_4 genes and evolutionary pattern in the C_4 metabolic phenotype. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 377–410. Academic Press, San Diego.

Mueller-Dombois, D., and Goldammer, J. B. (1990). Fire in tropical ecosystems and global environmental change: An introduction. In *Fire in the Tropical Biota—Ecosystem Processes and Global Challenges* (J. G. Goldammer, ed.), pp. 1–10. Springer-Verlag, Berlin.

Sage, R. F. (1995). Was low atmospheric CO_2 during the Pleistocene a limiting factor for the origin of agriculture? *Global Change Biol.* 1, 93–106.

Sage, R. F. (1999). Why C_4 photosynthesis? In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 3–16. Academic Press, San Diego.

Sage, R. F., and Pearcy, R. W. (2000). The physiological ecology of C_4 photosynthesis. In *Photosynthesis: Physiology and Metabolism* (R. C. Leegood, T. D. Sharkey, and S. von Caemmerer, eds.), pp. 497–532. Kluwer Academic, Dordrecht, Netherlands.

Sage, R. F., Li, M-R., and Monson, R. K. (1999a). The taxonomic distribution of C_4 photosynthesis. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 551–584. Academic Press, San Diego.

Sage, R. F., Wedin, D. A., and Li, M-R. (1999b). The biogeography of C_4 photosynthesis; Patterns and controlling factors. In *C_4 Plant Biology* (R. F. Sage and R. K. Monson, eds.), pp. 313–373. Academic Press, San Diego.

Smith, J. A. C., and Winter, K. (1996). Taxonomic distribution of Crassulacean acid metabolism. In *Crassulacean Acid Metabolism: Biochemistry, Ecophysiology and Evolution* (K. Winter and J. A. C. Smith, eds.), pp. 427–436. Springer, Berlin.

Soros, C. L., and Bruhl, J. J. (2000). Multiple evolutionary origins of C_4 photosynthesis in the Cyperaceae. In *Monocots—Systematics and Evolution. Vol 1 of the Preceedings of the Second International Conference on the Comparative Biology of Monocots* (K. L. Wilson and D. A. Morrison, eds.). CSIRO, Melbourne, in press.

Sponsel, L. E., Bailey, R. C., and Headland, T. N. (1996). Anthropological perspectives on the causes, consequences, and solutions of deforestation. In *Tropical Deforestation, the Human Dimension* (L. E. Sponsel, T. N. Headland, and R. C. Bailey, eds.), pp. 3–54. Columbia University Press, New York.

Thommason, J. R., Nelson, M. E., and Zakrzewski, R. J. (1986). A fossil grass (Gramineae: Chloridoideae) from the Miocene with Kranz anatomy. *Science* 233, 876–878.

Tidewell, W. D., and Nambudiri, E. M. V. (1989). *Tomlinsonia thomassonii*, gen. et sp. nov., a permineralized grass from the Upper Miocene Ricardo Formation, California. *Rev. Palaeobot. Palynol.* 60, 165–177.

CAPTIVE BREEDING AND REINTRODUCTION

Katherine Ralls* and Robin Meadows†
**Smithsonian Institution and †Freelance science writer*

GLOSSARY

allele One of two or more alternative forms of a gene.

effective population size The size of an ideal population that would have the same rate of increase in inbreeding or decrease in genetic diversity as the actual population under consideration. The effective population size is usually much less than the actual population size.

founder Wild-caught individual that contributes genetically to the captive population.

founder genome equivalent The number of equally contributing founders that would have produced the same genetic diversity found in an existing captive population if there had been no random loss of founder alleles.

genetic drift The variation of allele frequency from one generation to the next that occurs due to chance. Genetic drift leads to the loss of genetic variation in small populations due to the random loss of founder alleles during reproduction.

heterozygosity Average proportion of loci that are heterozygous (have two different alleles in an individual) in a population.

ideal population A hypothetical population widely used in population genetics theory. In this ideal population, the breeding sex ratio is equal, mating is random, generations do not overlap, selection and mutation do not occur, and the lifetime number of offspring produced by individual parents has a Poisson distribution.

mean kinship This value, calculated for every living member of a captive population, is the average kinship between that individual and all members of the population (including itself). Typically, living founders are excluded in the calculation of mean kinships. A population's average mean kinship is the average of the mean kinships of all the individuals in the population.

reintroduction Releasing individuals of a species into an area where that species no longer occurs in an effort to reestablish a wild population. Reintroduced individuals may be captured from a healthy wild population in another area or may be derived from a captive population if there are no healthy wild populations remaining.

studbook A list of all the living and dead individuals in a captive population that contains information on the mother, father, date of birth, location, and other topics for each individual.

Encyclopedia of Biodiversity, Volume 1

CAPTIVE BREEDING PROGRAMS enable zoos to exhibit many species of animals without capturing new individuals from the wild. Furthermore, the establishment of a captive breeding and reintroduction program may be the only hope for preserving a species that has been reduced to a very small number of individuals. Captive breeding programs are also used to support research and conservation education. A substantial research effort is often necessary to develop successful methods for raising a particular species in captivity and reintroducing it to the wild because these methods tend to be specific to a single species or a group of related species. However, methods for the demographic and genetic management of captive populations are similar across species. A captive breeding and reintroduction program should always be part of a larger conservation program for a species that also addresses the problems facing the species in the wild. Captive breeding and reintroduction programs require sustained long-term, adequately funded efforts and success is far from guaranteed.

I. NEED FOR CAPTIVE BREEDING AND REINTRODUCTION PROGRAMS

Captive breeding is the only choice for species that are extinct or nearly extinct in the wild. Nearly one-fourth of mammal species, 11% of bird species, and 25% of reptiles are threatened with extinction, according to the 1996 IUCN (World Conservation Union) *List of Threatened Animals*. Because most at-risk species are threatened by habitat loss and degradation, the most common conservation need is habitat protection and improvement. Captive breeding and reintroduction programs play a minor conservation role in comparison to protecting and improving habitat.

The need for captive breeding will undoubtedly increase because we are only just past the midpoint of the human population explosion, according to United Nations (UN) projections. The human population increased by a record-breaking 2 billion people in the past 25 years to a total of more than 6 billion, and the UN projects an increase of another 2 billion during the next 25 years with a total population of nearly 10.4 billion by 2100.

Due to limited space, staff, and funds, zoos will not be able to preserve populations of all animal species likely to become extinct in the wild. In the United States alone, the American Zoo Association (AZA) already maintains studbooks for 349 species. Of these, 118 species are managed under 89 interzoo captive breeding programs known as species survival plans (SSPs). The zoos participating in these plans move individuals of the species concerned to other zoos as necessary to meet genetic and demographic goals specified in the plan. Of the remaining 231 species, 105 are covered by less formal population management plans. AZA anticipates that at least an additional 136 studbooks and 18 SSPs will be organized within the next 3 years. At least one-third of the SSPs are involved with reintroduction efforts. Some species may be maintained in captivity for long periods without the possibility of reintroduction. For example, Père David's deer and the Mongolian wild horse survived in captivity many decades after their extinctions in the wild before the possibility of reintroducing them was even considered. Ultimately, a stable or decreasing human population with more sustainable patterns of resource use may allow the reintroduction of some species that have been preserved only in captivity.

II. GENERAL ASPECTS OF CAPTIVE BREEDING AND REINTRODUCTION PROGRAMS

The major goal of most captive breeding programs is to develop self-sustaining captive and/or wild populations. Even if a species is never reintroduced, a successful captive breeding program will supply zoos with animals to exhibit, thereby minimizing the need to collect them from the wild. Captive breeding programs also have considerable educational value because they are used to inform zoo visitors of the value of conserving biodiversity and to increase public interest in conservation issues. Animals maintained in captive breeding programs also support a variety of research programs. For example, some aspects of human medicine, human evolution, and the distinctiveness of the human genome require comparison of human DNA with chimpanzee and gorilla DNA, samples of which are normally obtained from captive populations.

Several organizations provide help with captive breeding and reintroduction efforts. The AZA coordinates captive breeding efforts in the United States and similar organizations exist in other countries. The AZA maintains two committees that give general advice on population management and reintroduction: the Small Population Management Advisory Group and the Reintroduction Advisory Group. The IUCN maintains similar committees of international experts: the Conservation Breeding Specialist Group and the Reintroduction Specialist Group. Both the AZA and the IUCN have other committees that focus on particular groups of species, such as carnivores or primates, and then identify the species within these groups that would most benefit from captive breeding and/or reintroduction. The International Species Information System maintains a database of captive animals in zoos throughout the world and develops and distributes software for maintaining records on captive populations, including studbooks, and for facilitating genetic and demographic management of these populations.

III. DIFFICULTIES WITH CAPTIVE BREEDING PROGRAMS

A. Failure to Breed in Captivity

Not all species breed well in captivity. The species that are easiest to breed in captivity are those that have management requirements similar to those for domestic animals or for other species with which zoos have had considerable experience. For example, zoos suddenly faced with the challenge of breeding California condors experienced little difficulty because they had been breeding Andean condors for many years.

However, species with which zoos have had little prior experience may breed very poorly at first until zoos develop appropriate husbandry techniques. Thus, because husbandry techniques tend to be species specific, new captive breeding programs often require substantial research programs on behavior, reproductive biology, nutrition, genetics, or disease. Research on closely related species is also often helpful.

Poor reproduction in captivity is often due to behavioral problems caused by inadequate husbandry techniques. Because different zoos often have different degrees of success in breeding a particular species, important insights can often be gained by comparing the behavior and reproductive success of individuals kept under different conditions at different zoos. For example, a collaborative cross-zoo study indicated that more submissive male black rhinoceroses tended to sire more young and that reproductive success increased when the species was kept in larger cages. Space limitation or crowding may increase aggressiveness in male black rhinoceroses and therefore may reduce reproductive success. This suggests that dominant, aggressive males should be housed in zoos with large enclosures.

More invasive research work can also make important contributions. For example, studies on black-footed ferret reproductive biology have improved captive breeding techniques for this species. Until recently, only about 58% of mated black-footed ferret females produced young. Many of the other females became pseudopregnant, which means they underwent the hormonal changes typical of pregnancy without actually being pregnant. Now researchers know that yearling males do not produce sperm until 4–6 weeks later than older males. Many of the pseudopregnancies were the result of breeding these yearling males. Ensuring that males were actually producing sperm before using them for breeding has reduced the number of pseudopregnant females by about 20%.

B. Genetic and Behavioral Adaptation to Captivity

A captive population's risk of extinction is increased by inbreeding and loss of genetic variation, and inbred animals with little genetic variation are less likely to survive when reintroduced to the wild than more outbred individuals. However, both inbreeding and loss of genetic variation are unavoidable in small, closed populations because all individuals in the population eventually become related to each other. A randomly mating small population loses genetic variation (heterozygosity) at a rate equal to

$$\frac{1}{2N_e}$$

each generation, where N_e is the effective population size.

Captive populations are routinely managed to minimize inbreeding and loss of genetic variation. Another threat to captive populations is that selective pressures in captive habitats are different than those in wild habitats; thus, captive populations tend to adapt to captivity, which can make it more difficult to reestablish a population in the wild.

Learned behavioral traits can degenerate in captivity more rapidly than genetic diversity. Traits that may degenerate in the captive environment include foraging skills, detection and avoidance of predators, and fear of humans. When captive-bred and wild-born individuals have been experimentally released in the same location, the captive-bred individuals have tended to survive for shorter periods due to lack of appropriate behaviors.

IV. GENETIC AND DEMOGRAPHIC MANAGEMENT OF CAPTIVE POPULATIONS

In contrast to husbandry methods, genetic and demographic management methods are similar for all captive populations. Genetic and demographic management of captive populations focuses on maintaining genetic diversity in order to minimize undesirable genetic changes due to selection in the captive environment, avoid deleterious effects of inbreeding depression, and maintain future options for genetic management.

A. Starting the Program

Ideally, the first step in the development of a captive breeding and reintroduction program is agreement among all concerned parties, such as agency personnel, nongovernmental conservation groups, and outside scientific advisers, that such a program would benefit a particular species. This step may be difficult because some people oppose captive breeding in general and because there are no precise guidelines for when to begin a captive breeding program. However, the IUCN recommends starting a captive population long before the wild population reaches the critical state in order to increase the probability that the captive breeding program will be a success.

Once in place, captive breeding programs have three phases. In the founding phase, the captive population is started. In the growth phase, the population rapidly increases to the final "target" population size specified by its managers. In the carrying capacity phase, the population is maintained at its target size and excess individuals may be reintroduced into the wild (Fig. 1). Management concerns change as the captive population progresses through these phases.

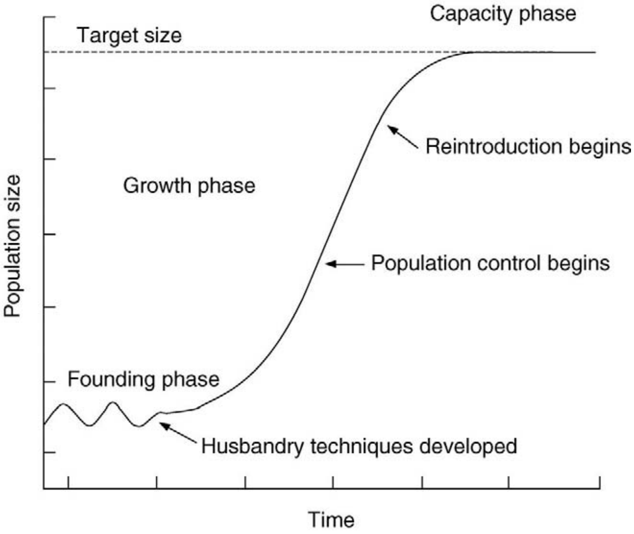


FIGURE 1 The development of a captive breeding and reintroduction program from the founding to the capacity phase.

B. Management Concerns during the Founding Phase

The main management concerns during the founding phase are removing individuals for the captive population with minimal impact on the wild population, acquiring enough founders from the wild to achieve genetic goals, getting the species to breed reliably in captivity, and setting general goals and plans for the captive population. Ways to capture animals for the captive population with minimal impact on the wild population include removing eggs from nests, using orphaned or injured animals, and capturing dispersing juveniles. Many birds (e.g., condors) will usually lay another egg to replace an egg that has been removed, and dispersing juvenile mammals often have high mortality rates in the wild. If the species is one that zoos do not know how to be breed reliably in captivity, it is best to solve husbandry problems with only a few wild individuals or even animals of a closely related species.

Genetic goals for a captive population are usually specified in terms of the proportion of genetic variation (measured as heterozygosity) to be maintained for a specified time. A common goal is to maintain 90% of the genetic diversity of the source population for 100 years. However, some programs use other time frames. For example, the Guam rail and black-footed ferret programs are using the goal of "90% for 50 years" because of the short generation times for these species (Table I) and plans for the rapid establishment of several wild populations.

Once a genetic goal has been set, population genetics theory enables calculation of the number of founders needed for the captive population (the number of wild animals that must be captured and successfully bred) and the target population size (the number of individuals that must be maintained in captivity during the planning period). Planning to retain a higher proportion of genetic variation usually increases the target population size. For example, maintaining 90% of the ferret genetic variation for 50 years requires a population of only 500 individuals, whereas maintaining 92% would require a population of 2700. Increasing the length of

TABLE I

The Goals and Number of Founders of Captive Breeding Programs with Reintroduction Components[a]

	Species			
	California condor[b]	Black-footed ferret	Guam rail	Golden lion tamarin
Heterozygosity goal (%)	90	90	90	90
Length of program (years)	200	50	50	200
Number of generations	10	20	22	33
Target population size	150	500	150	550
Number wild-caught	14	18	21	69[e]
Number of contributing founders[c]	13	10	13	45
Founder genome equivalents[d]	8	5	5	12

[a] Species are listed in order of increasing number of generations in the program length (from Ralls and Ballou, 1992).

[b] Heterozygosity goal, program length, and target population size have not been officially adopted by program managers; other data from Kieler (1991).

[c] Founders with currently living descendants.

[d] The number of theoretically ideal founders taking into consideration loss of genetic diversity in the current captive population (Lacy, 1989).

[e] Includes the number of wild-caught tamarins acquired after the captive program was initiated in 1981 in addition to the number of founders and wild-caught individuals alive at the initiation of the program.

the planning period has a similar effect. For example, maintaining 90% of the ferret genetic variation for 100 instead of 50 years would increase the target population size from 500 to 1300 individuals.

Increasing the number of founders reduces the size of the target population needed to reach a particular goal. Twenty to 30 unrelated individuals are generally a sufficient number of founders. Unfortunately, many existing captive breeding programs were begun after it was already too late to acquire this many founders. For example, the ferret population had only 10 founders. If it had been possible to obtain 25 founders, the target population size could have been reduced from 500 to 200 individuals. Although a small number of founders reduces the probability that a captive breeding program will be successful, it does not doom it to failure. Thus, the lack of an ideal number of founders does not justify abandoning or failing to initiate a captive breeding program.

The target population size also depends on the species' rate of reproduction and generation length. A smaller target population will be required to reach the genetic goal if the species can grow more rapidly each generation or if it has a long generation time (because genetic variation is lost due to genetic drift each time individuals reproduce). The target population size may also be limited by practical considerations, such as the number of spaces available in zoos. Fewer zoos may be willing to participate in the program if the species is not attractive as an exhibit. Thus, the target population size may be a compromise between genetic and demographic factors and the limited resources available.

Once husbandry problems have been solved and the species is breeding well in captivity, the rest of the founders should be obtained as soon as possible. Unfortunately, the number of animals that must be captured from the wild is usually greater than the number of founders needed. Wild-caught animals may be related or fail to breed, or their descendants may fail to reproduce. For example, although 25 wild black-footed ferrets were captured, some died of distemper, some were known to be parents and offspring, and some failed to reproduce. Although ideally each founder will contribute an equal number of offspring to the captive population, those ferrets that did reproduce did so unequally, skewing their genetic contributions to the captive population. Ultimately, the ferret population was founded by the theoretical genetic equivalent of only 5 ferrets; that is, 5 founder genome equivalents.

C. Management Concerns during the Growth Phase

Management efforts during the growth phase center on getting the population to increase as rapidly as possible. Rapid growth has two benefits: It increases the captive population's chances of survival and it retains as much of the founders' genetic diversity as possible. Small captive populations are at higher risk of extinction due to many factors, including random demographic events (such as a succession of male births), inbreeding depression, and unpredictable events that can kill numerous individuals such as diseases, fires, hurricanes, and other catastrophes.

The standard SSP breeding strategy used in the United States is designed to maximize the retention of genetic diversity. This is accomplished by minimizing mean kinship among the members of the captive population (Ballou and Lacy, 1995). Breeding pairs are formed based on mean kinship, beginning with the individuals with lowest mean kinships, until the desired number of pairs is attained. Efforts are also made to avoid mating closely related individuals when forming new pairs. During the growth phase, this strategy is modified slightly to choose new pairs to minimize mean kinship as much as possible but breed all individuals in the population.

Although managers attempt to minimize mean kinship and inbreeding during the growth phase, rapid population growth takes priority over genetic concerns, particularly when the population is very small and the risk of extinction outweighs the risk of a few less-than-ideal matings. For instance, if a female rejects the genetically ideal mate, she may be allowed to mate with another male she prefers. This problem has occurred several times in the California condor population.

At some point during the growth phase, the captive population usually is divided into subpopulations housed in different breeding facilities. This reduces the risk that a catastrophe such as disease or fire will decimate the entire captive population. To ensure that each subpopulation is as genetically diverse as possible, each should have individuals descended from each of the founders.

D. Management Concerns during the Capacity Phase

Once the population has reached the target size, relatively few offspring may be needed each year to maintain it at that level. Thus, genetic concerns become more important and managers select which individuals to breed.

The number of offspring needed to maintain the captive population can be calculated by standard demographic techniques. Any “extra” offspring can be used for reintroduction. If there are more offspring than are needed for reintroduction, managers can prevent some adults from breeding either by using contraceptives or by housing them individually or in same-sex groups. For example, the captive golden lion tamarin population is now at its target size of about 550 individuals. Because tamarins often have twins or even triplets, these 550 individuals could potentially produce more than 400 offspring per year. However, only about 80 per year are needed to maintain the captive population and augment wild ones (the number that can be reintroduced to Brazil is limited by funding and habitat availability). Consequently, each year approximately 100 female tamarins are prevented from breeding by being housed with another female or a nonreproductive male or by using hormonal contraceptive implants.

There are two general strategies for producing the individuals to be reintroduced. If the date of a reintroduction effort can be scheduled well in advance and the species has a predictable breeding pattern (such as breeding once a year during the spring), males and females can be paired up for the specific purpose of producing excess young for that particular reintroduction. This approach has been used for Guam rails. However, if the date of a reintroduction effort is difficult to predict in advance (this may occur due to difficulties with funding or permits), animals for reintroduction can be selected from the existing population and breeding pairs can be set up to replace the reintroduced individuals.

In the early stages of a reintroduction program, reintroduction techniques are still being refined and mortality may be high. Thus, initially the most genetically expendable individuals are usually released. Later, emphasis will gradually shift to choosing individuals that are not closely related to the individuals already present in the wild population. This maximizes the genetic diversity of the wild population. The final genetic goal is to make the wild population as genetically diverse as the captive population.

V. REINTRODUCING CAPTIVE-BRED ANIMALS TO THE WILD

Ideally, the goals of all captive breeding plans would include reintroduction back to the wild. However, some species may be impossible to reintroduce due to lack of habitat or other problems. Furthermore, some species will be easier to reintroduce than others. Mark Stanley Price, former chairman of the IUCN Reintroduction Specialist Group, suggests that the species that may be easiest to reintroduce include large species with few predators, those living in herds or other social groups, those that tend to explore new habitat, and those that can tolerate a wide range of habitat conditions.

TABLE II

The Elements of a Successful Captive Breeding and Reintroduction Program[a]

- Captive population
  - Ongoing research in behavior, genetics, physiology, nutrition, reproduction, and pathology
  - Genetic and demographic management of the population
  - Self-sustaining viable captive population
- Field studies
  - Regular censuses of the size, distribution, and genetics of the wild population
  - Behavioral ecology studies (home range size, movements, habitat preferences, social organizations, mating system, feeding, and antipredator adaptions)
  - Locating existing suitable habitat containing critical resources for reintroduction
- Habitat preservation and management
  - Protection of habitat from degradation and exploitation
  - Restoration and management of degraded habitats
  - Increase in or maintenance of the number of preservation areas
- Conservation education for long-term support
  - Professional training through academic studies, workshops, internships, courses, and fellowships
  - Determining the most appropriate public relations and educational strategies through surveys
  - Public relations educational efforts using appropriate mass media (e.g., television, radio, magazines, and newspapers)
  - Local community education, both formal and informal
- Preparation and reintroductions of animals
  - Choice of candidates and assessment of their characters for retrospective correlation with postrelease survival
  - Training in survival techniques, including foraging and feeding, antipredator tactics, locomotion, and orientation
  - Adaption to local conditions at release site (food, climate and temperature, and disease)
  - Release and long-term monitoring to evaluate causes of death and basis for survival

[a] From Kleinman (1989).

The elements of a successful reintroduction program involving captive-bred mammals are shown in Table II. These elements include research on both the captive and wild populations, habitat preservation and management, conservation education to ensure long-term support of the program, and careful management and monitoring of the reintroduced individuals. The IUCN also offers a set of guidelines that discusses the biological, socioeconomic, and legal requirements for successful reintroduction.

Most important, reintroduction is a realistic goal only when habitat protection is an integral part of the species' overall conservation plan. A species should not be reintroduced unless the factors that led to its decline in the wild have been identified and eliminated—or at least greatly reduced—and suitable legally protected habitat exists. In addition, the release site should be within the species' historic range. Occasionally, however, a species must be "reintroduced" into areas of suitable habitat outside of its historic range. For example, the Guam rail is being reintroduced to the nearby island of Rota because the nonnative brown tree snake, which led to the bird's extinction in the wild, has invaded its entire historic range on Guam. Similarly, a variety of species, including birds, reptiles, and invertebrates, from New Zealand and Australia are being reintroduced to offshore islands that are free of the nonnative predators that led to their extinction on the mainland.

There are many other factors to consider when reintroducing captive-bred individuals. For instance, the release of captive-bred animals can spread diseases to an existing wild population or disrupt its social organization. Thus, it is advisable to screen captive-bred individuals for diseases prior to release and to release them in habitat that has no wild individuals.

When reintroduced to the wild, captive-bred individuals are likely to suffer high mortality rates due to inappropriate behavior. For example, they may have difficulty finding enough food or fail to avoid predators. It has proven very difficult to help orphaned sea otter pups raised in captivity develop appropriate foraging skills and teach them to avoid humans. Captive-bred condors develop appropriate foraging skills fairly easily but often fail to avoid humans and human structures. Substantial research programs are often needed to develop husbandry and reintroduction techniques that will promote behaviors needed for survival in the wild.

The conditions under which captive-bred individuals are raised can be critical. The development of appropriate survival skills may require a skilled parent or a particular stimulus at some critical period during development. For example, adult ferrets prefer eating whatever they were fed when they were 2 or 3 months old, which is when ferrets develop their permanent teeth. Therefore, at the age of 2 or 3 months, captive ferrets should be fed prairie dogs, their exclusive prey in the wild. Methods of reintroduction may also require research. For example, should the animals be released as social groups or as individuals? Should they be fed after they are released and, if so, for how long? The answers to such questions depend on the particular species being

TABLE III

Checklist for Deciding Whether or Not Appropriate Conditions Exist for Beginning a Reintroduction Program Applied to Three Species of Lion Tamarins[a]

	Lion tamarin species		
	Golden	Golden-headed	Black
Condition of species			
1. Need to augment wild popn.	Yes	No	Yes(?)
2. Available stock	Yes	Yes	No
3. No jeopardy to wild popn.	?	?	?
Environmental conditions			
4. Causes of decline removed	?	No	No
5. Sufficient protected habitat	Yes(?)	No	Yes
6. Unsaturated habitat	Yes	Yes(?)	?
Biopolitical conditions			
7. No negative impact for locals	No	?	?
8. Community support exists	5	2	4
9. GOs/NGOs supportive/involved	Yes	Yes	Yes
10. Conformity with all laws/regulations	Yes	?	?
Biological and other resources			
11. Reintroduction technology known/in development	4	3	3
12. Knowledge of species' biology	5	1.5	3
13. Sufficient resources exist for program	Yes	No	No
Recommended reintroduction/translocation?	Yes	No	No

[a] Adapted from Kleiman (1990).

reintroduced. Reintroduction programs using captive-bred individuals are usually expensive, lengthy, complex, and difficult. Thus, the decision to begin such a program should not be made lightly. A short checklist of the major factors that should be considered when deciding whether or not to reintroduce a species is illustrated in Table III with respect to three species of lion tamarins. Answers to the questions in the checklist indicate that reintroduction is appropriate for golden lion tamarins but not the other two species because the causes of their decline have not been eliminated and funds to support a reintroduction program are not available.

There are no generally accepted guidelines for declaring the success of a reintroduction effort. Beck and colleagues (1994) suggested two possible criteria: if 500 wild individuals survived without human support or if a formal population viability analysis predicted that the population would be self-sustaining. By these criteria, they judged that only 11% of 145 animal reintroductions were successful.

However, many of the reintroduction programs they considered are still in progress and it is too early to evaluate their success. For example, techniques for breeding and reintroducing black-footed ferrets have greatly improved and are still improving. Since 1986, the black-footed ferret program has released 873 captive-bred ferrets at five sites. The combined wild population is probably larger than the captive population, which is maintained at 240 breeding individuals. In 1998, more than 30 litters containing more than 100 pups were born in the wild.

Successful captive breeding and reintroduction programs require sustained long-term, adequately funded efforts. Research can solve many problems involved in successfully breeding a species in captivity and reintroducing it to the wild. However, the ultimate success of many programs, such as that for the ferret, will depend on whether or not we are able to preserve enough suitable habitat to sustain viable wild populations of the species.

See Also the Following Articles

BREEDING OF ANIMALS • BREEDING OF PLANTS • CONSERVATION EFFORTS, CONTEMPORARY •

ENDANGERED BIRDS • ENDANGERED MAMMALS • ENDANGERED REPTILES AND AMPHIBIANS • ZOOS AND ZOOLOGICAL PARKS

Bibliography

Ballou, J. D., and Lacy, R. C. (1995). Identifying genetically important individuals for management of genetic diversity in captive populations. In *Population Management for Survival and Recovery* (Ballou, J. D., Gilpin, M., and Foose, T. Eds.), pp. 76–111. Columbia Univ. Press, New York.

Endangered Species UPDATE (1990). Captive breeding and reintroduction: A strategy for preserving endangered species? Vol. **8**, No. 1, pp. 1–88.

Frankham, R. (1995). Conservation genetics. *Annu. Rev. Genet.* **29**, 305–327.

Kleiman, D. G., Allen, M. E., Thompson, K. V., and Lumpkin, S. (Eds.) (1996). *Wild Mammals in Captivity*. Univ. of Chicago Press, Chicago.

Olney, P. J. S., Mace, G. M., and Feistner, A. G. C. (Eds.) (1994). *Creative Conservation: Interactive Management of Wild and Captive Populations*. Chapman & Hall, London. [Includes Beck *et al.* (1994) article mentioned in text]

Serena, M. (Ed.) (1995). *Reintroduction Biology of Australian and New Zealand Fauna*. Surrey, Beatty, Norton, New South Wales, Australia.

CARBON CYCLE

John Grace
University of Edinburgh

- I. General Features
- II. Response of the Oceanic Sink to Climate Change
- III. Response of the Terrestrial Sink to Climate Change
- IV. History of the Carbon Cycle
- V. Managing the Carbon Cycle
- VI. Will the Sinks Increase, Remain the Same, or Decrease?

GLOSSARY

autotrophic respiration The respiration of photosynthetic organisms.

biomass The mass of biological material after removal of water by oven-drying at 70–100°C, often expressed as mass per area of ground surface.

BP Before the present, usually measured in billions of years (Ga) or millions of years (Ma).

El Niño–Southern Oscillation Global climatological phenomenon, occurring every few years, involving reversal of trade winds, warming of parts of the ocean, and consequent widespread disturbances to the climate of several regions.

global circulation models (GCMs) Computer models of atmospheric circulation patterns and surface energy balance used for weather forecasting and to predict climates at regional and global scales. GCMs are so large that they must be run on supercomputers.

heterotrophic respiration The respiration of all those life forms that feed on photosynthetic organisms, including bacteria, fungi, and animals.

isotope Many elements exist in several forms, called isotopes. Carbon has eight isotopes, of which ^{12}C forms 98.9% of naturally occurring CO_2 and ^{13}C forms 1.1%. Other carbon isotopes are unstable, although ^{14}C has a half-life as long as 5730 years.

isotopic discrimination The tendency of chemical and physical reactions to "prefer" one isotope against another. The enzyme ribulose bisphosphate carboxylase oxygenase (Rubisco), which is responsible for capturing CO_2 in photosynthesis, shows strong discrimination for $^{12}CO_2$ against $^{13}CO_2$.

lignin Heterogeneous carbon polymer associated with cellulose to form wood; resistant to decay by bacteria, and broken down by some fungi.

Q_{10} Mathematical index expressing the effect of temperature on respiration or decomposition: $Q_{10} = 2$ means that the rate doubles for a 10°C increase in temperature.

sequestration Term used to describe the uptake of carbon from a dilute source (the atmosphere) to a concentrated form (biomass).

CARBON IN THE FORM OF CO_2 is exchanged by natural processes between the atmosphere and the land, and between the atmosphere and the ocean. The natural processes involved are photosynthesis, respiration, and dissolution. The gains and losses may be viewed as a cyclic process, generally called the carbon cycle.

The cycle is not in equilibrium. As a result of burning of fossil fuels and replacement of forest by farmland there is currently an anthropogenic input to the atmosphere of about 8 billion tonnes of carbon each year as CO_2. Of this, only about 3 billion tonnes of carbon appears in the atmosphere and contributes to global warming via the greenhouse effect. The remainder dissolves in the ocean, or is taken up by photosynthesis and stored as biomass or organic matter in the soil.

Photosynthesis is increased at elevated CO_2, so a negative feedback process may be operating whereby elevated CO_2 increases the capacity of the terrestrial ecosystems to act as a carbon sink. Nitrogen compounds emitted from terrestrial sources, especially motor engines and farmland, are deposited worldwide and may be enhancing the sink effect. But the long term future of the sink is not assured. Global warming is likely to increase microbial decomposition of soil organic matter, causing an increased transfer of C from soil to atmosphere, hence reducing the sink.

There is now widespread realization, by world governments as well as scientists, that anthropogenic emissions of CO_2 and other trace gases are increasing the greenhouse effect and causing global warming. Consequently, there is renewed interest in the carbon cycle, and in managing both the terrestrial and ocean sinks to reduce the rate at which CO_2 increases in the atmosphere.

I. GENERAL FEATURES

The carbon cycle is a natural biogeochemical cycle in which carbon as CO_2 is transferred from the atmosphere to the land and ocean, where it resides in another form before returning to the atmosphere as CO_2. The principal processes involved in transfer from the atmosphere are the dissolution of CO_2 in the oceans and the uptake of CO_2 by the photosynthesis of green plants. The processes involved in return to the atmosphere are the release of CO_2 from the ocean in regions in which the surface of the ocean has become saturated with CO_2 and the oxidation of organic matter by respiration or fire, which essentially reverses the photosynthetic process:

$$\begin{array}{c}\text{photosynthesis} \rightarrow \\ 6CO_2 + 12H_2O \Leftrightarrow C_6H_{12}O_6 + 6O_2 + 6H_2O \\ \text{respiration and} \\ \leftarrow \text{fire.}\end{array} \tag{1}$$

In photosynthesis, water is split and the 12 oxygen atoms produce six molecules of oxygen gas. Carbon dioxide is reduced to glucose ($C_6H_{12}O_6$), from which other biochemical constituents of biomass are synthesized.

The carbon cycle can be envisioned as a set of fluxes between major pools (Fig. 1). The pools differ in magnitude, and in the average time a carbon atom resides in them, and therefore the dynamic behavior of the system as a whole is likely to be complex. For example, an "average" carbon atom can be expected to reside in the atmosphere for about 5 years, whereas in the ocean the corresponding residence time will be hundreds of years.

An understanding of the carbon cycle is fundamental to our understanding of life because all biomass is carbon based. The overall carbon content of dry biomass is in the range 45–55%. The carbon cycle has become a subject of intense research in recent years, since the realization that anthropogenic emissions of CO_2 are large enough to force the cycle into a state of disequilibrium whereby the concentration of the gas in the atmosphere is increasing. Because carbon dioxide is the most important of the several "greenhouse gases" that absorb infrared radiation emitted from the planetary surface, its increase in concentration is capable of causing additional global warming. Indeed, there is overwhelming evidence that it is already doing so. The Framework Convention on Climate Change (United Nations, 1992)

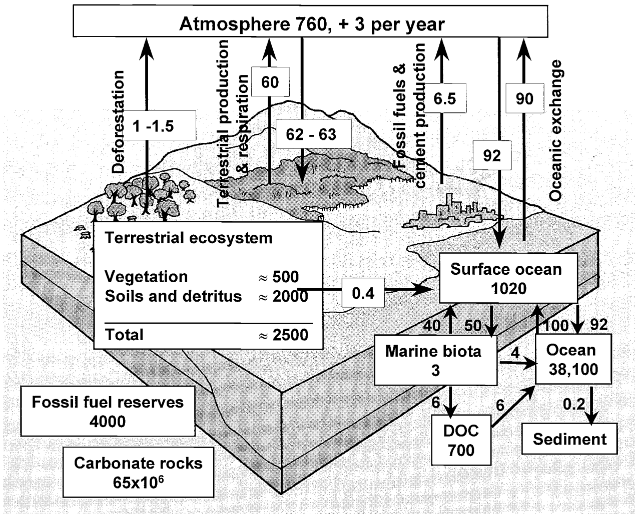


FIGURE 1 The global carbon cycle during the 1990s. The carbon stocks (shown in boxes) are in billions of tons of carbon (GtC); the carbon fluxes (arrows) are in GtC per annum. Note that both the terrestrial ecosystem and the ocean are net absorbers of CO_2. The atmospheric stock is increasing by 3 GtC per annum. DOC, dissolved organic carbon.

TABLE I

Definition of Units Used in This Article

Quantity	Units	Abbreviation	Relation to alternative units
Area	Hectares	ha	1 ha = 10^4 m^2 = 0.01 km^2
Carbon stocks	Gigatons of carbon	GtC	1 GtC = 1 × 10^9 tC = 1 PgC = 10^{15} gC = 3.66 GtCO$_2$
Carbon fluxes (global)	Gigatons of carbon per annum	GtC a^{-1}	
Carbon fluxes (ecosystem)	Gigatons of carbon per annum per hectare	tC ha^{-1} a^{-1}	1 tC ha^{-1} a^{-1} = 100 g m^{-2} a^{-1}
Concentration of CO$_2$	Parts per million by volume	ppm	1ppm = 1 μmol C (mol air)$^{-1}$

states an ultimate aim: "to achieve stabilization of greenhouse gas concentrations . . . at a level that would prevent dangerous anthropogenic interference with the climate system."

In the early 1990s, it was noted that knowledge of the carbon cycle is still incomplete. In particular, it was shown that the annual anthropogenic emissions of carbon as CO$_2$ exceeded the sum of the annual increase observed in the atmosphere plus that dissolving in the oceans. In other words, it was not possible to account for all the emitted CO$_2$. It was postulated that there must be a substantial "missing sink" for CO$_2$, amounting to about 2 billion tons of carbon (2 GtC). The sink was presumed to be the net photosynthetic uptake of the terrestrial vegetation. Therefore, in the 1990s there was a major research effort to resolve uncertainties in the location and magnitude of sinks. The challenge now is to predict the behavior of the carbon cycle over the next few decades, as CO$_2$ and temperature continue to rise. The goal is to manage the carbon cycle through using scientific knowledge to develop policy.

The cycle shown in Fig. 1 is modified from Schimel (1995), updated to take into account new data on the rate of fossil fuel burning and reducing slightly the deforestation flux according to recent estimates. Units are in billions of tons of carbon, GtC (see Table I for conversion factors). In general, there is considerable uncertainty regarding many of the data on carbon stocks, largely the result of the heterogeneous nature of the earth's aquatic and terrestrial surface and the difficulty in comprehensively sampling the spatial variability.

Gases other than carbon dioxide are involved in the global carbon cycle, but they are quantitatively less important. Methane (CH$_4$) and isoprene (C$_5$H$_8$) both have global emission rates estimated as about 0.3–0.5 GtC a^{-1}, much less than the respiratory fluxes of carbon dioxide. Carbon monoxide is formed during combustion, and also by oxidation of methane, but it reacts with the OH radical in the atmosphere to form carbon dioxide. The total production of carbon monoxide from all sources is estimated as 2.4 GtC a^{-1} but its lifetime is only a few weeks before oxidation to CO$_2$.

A. Atmospheric Analysis

Some of the best information about the carbon cycle comes from the analysis of the CO$_2$ concentration in the atmosphere, pioneered in the 1950s by Keeling, who established a CO$_2$ observatory at Mauna Loa in Hawaii and first demonstrated the upward trend in the CO$_2$ concentration (Fig. 2). Superimposed on the trend, there is an annual cycle whereby the concentration decreases during summer in the Northern Hemisphere

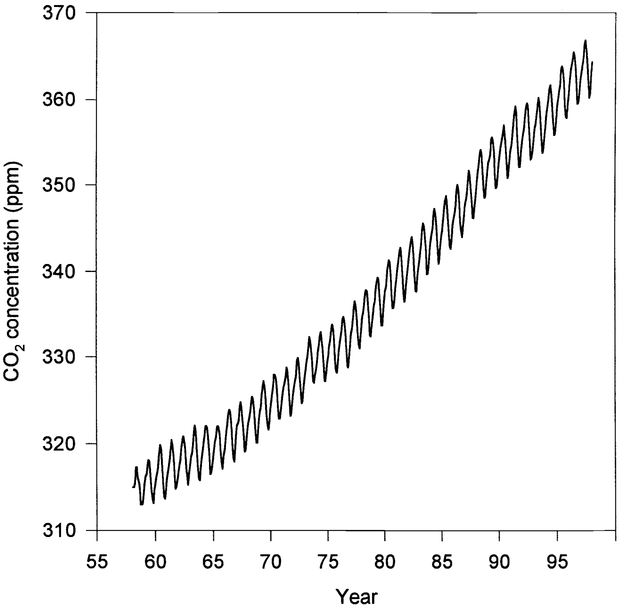


FIGURE 2 The increase in CO$_2$ concentration (parts per million by volume) as measured over 40 years at Mauna Loa in Hawaii.

and increases in the winter, with a minimum in October and a maximum in May. This summer decline is attributable to strong summertime uptake by photosynthesis of terrestrial vegetation (in the Southern Hemisphere, there is much less land and so a corresponding photosynthetic signal is not evident during the southern summer). Now, there is a network of remote stations whereby air samples are regularly taken in glass flasks and sent to a common laboratory for analysis. An important component of the analysis is the isotopic signal of CO_2. This enables us to distinguish between ocean uptake and the photosynthetic uptake by C_3 plants because the latter discriminates against ^{13}C, whereas the former does not. Recently, it has become technically possible to detect small changes between oxygen and nitrogen concentrations. This also provides a signal of photosynthesis because photosynthesis releases one molecule of O_2 for every CO_2 taken up, whereas dissolution in the ocean has no influence on O_2. In fact, just as the CO_2 concentration is increasing by a few parts per million (ppm) each year so also the O_2 concentration is decreasing. Fortunately, this is not cause for alarm because the O_2 concentration is very high (about 210,000 ppm). Ultimately, this technique of measuring changes in O_2 may prove to be the most sensitive method of detecting trends in photosynthesis at a global scale. Currently, most of the inferences have been made from CO_2 concentrations and isotopic fractions of ^{13}C and ^{12}C. Using data from the flask network coupled with knowledge of the anthropogenic emissions and atmospheric circulation, it is possible to calculate the latitudinal distribution of the terrestrial and oceanic carbon sink. Currently, there is not complete agreement between different groups of workers on the point of detail because each group uses its own approach to the calculation, but three conclusions emerge:

1. There is a large northern net sink of carbon, associated with uptake by the terrestrial vegetation and the ocean, usually estimated as 1 or 2 Gt of carbon per year.
2. In the equatorial latitudes there is a small net sink, but because deforestation accounts for an efflux of about 1.5 Gt of carbon per year there must be a biotic sink of opposite sign and about the same magnitude.
3. The pattern of sink distribution is not the same each year, being influenced by climatic phenomena and possibly by major volcanic eruptions (droughts associated with the El Niño–Southern Oscillation have been implicated as the main influence).

Overall, we can conclude from these studies that the net terrestrial sink is 2 or 3 GtC of carbon per year, and the ocean sink is likely to be about 2 GtC a^{-1}. In addition to the flask measurements mentioned previously, there is independent evidence for terrestrial carbon sinks. For example, temperate forests in Europe and North America are growing faster than ever before, and in the equatorial region it has been found that undisturbed, mature forests are accumulating carbon.

B. Ocean Analysis

The surface of the planet is largely covered by ocean. There are 361×10^6 km² of ocean compared with 149×10^6 km² of land. Carbon dioxide is readily soluble in water. At 10°C the solubility is 1.16 m³ of gas/m³ of water, but this decreases to 0.85 at 20°C. In solution, there is an equilibrium mixture of carbonic acid and bicarbonate and carbonate ions, which make up the fraction "dissolved inorganic carbon" in Fig. 1:

$$CO_2 + H_2O \Leftrightarrow H_2CO_3 \Leftrightarrow H^+ + HCO_3^- \Leftrightarrow 2H^+ + CO_3^{--} \quad (2)$$

The proportions of each species depend on pH. At high pH the reactions shift to the right. At a pH of approximately 8, in the ocean most of the carbon is in the form of bicarbonate. In certain conditions of very high pH, carbonate predominates and calcium or magnesium carbonate are precipitated from the solution. Most of the world's limestone since Cambrian times consists of shells and exoskeletons of marine life. In present-day seas the main groups of organisms responsible are mollusks, corals, echinoderms, foraminifera, and calcareous algae.

Marine ecosystems play a role in the carbon cycle through the so-called "biological pump" acting as follows. Organisms occupy the well-mixed surface layers of the ocean and photosynthesize and grow at a rate which varies according to the nutritional state of the ocean. Dead biota and feces fall through the water column, some of them reaching the seabed, thus removing carbon from the surface layers and hence reducing the partial pressure of CO_2 there. This enables uptake of new CO_2 from the atmosphere. Thus, the ocean's sink strength is increased by biological activity. As the particles sink, decomposition occurs, with the release of inorganic carbon and nutrients. Attempts to directly measure the carbon transferred to the seabed in this particle "rain" by means of traps suggest that the downward flux of particulates is very low (Fig. 1). At great depth, the high pressure enables more CO_2 to be dissolved in the water than would otherwise be the case.

The global ocean has currents that operate on various scales of time and space. Cold water subsides in polar regions, setting up a circulation. This tendency is enhanced by the fact that polar water is more salty and therefore more dense in the winter, as the fresh water is "frozen out" and added to the polar ice caps. A thermohaline circulation pattern is thus set up, with water subsiding in the polar regions and upwelling in tropical waters (Fig. 3). Patterns and rates of circulation have been obtained from studying the transport of material which entered the ocean as fallout from nuclear weapons tests in the early 1960s. The circulation is very slow compared with atmospheric circulation. In the equatorial zone the upwelling brings nutrient- and carbon-rich water to the surface and outgassing of CO_2 occurs.

Estimates of the sink strength of the ocean are made using knowledge of the CO_2 partial pressure in the surface waters. The rate of uptake or loss of CO_2 is proportional to the difference in partial pressure between the water and the air and to the exchange coefficient which depends on wind speed. Tens of thousands of such measurements are made annually. In high latitudes the surface waters are less than saturated with CO_2, so CO_2 dissolves from the atmosphere. In equatorial regions many waters are saturated with CO_2 and so there is an efflux to the atmosphere. On average, the ocean's uptake is calculated as about 2 GtC per year. A recent CO_2 flux map indicates strong sinks in the north Atlantic and Pacific and a source area in the mid-Pacific corresponding to an efflux of 0.5–1 GtC per year. There is likely to be considerable interannual variability in the sink strength of the ocean as a result of variations in currents, which influence sea surface temperatures and may influence the extent to which CO_2-rich water is brought to the surface. For example, the efflux from the equatorial waters is increased during El Niño years, when the surface temperature of the Pacific increases.

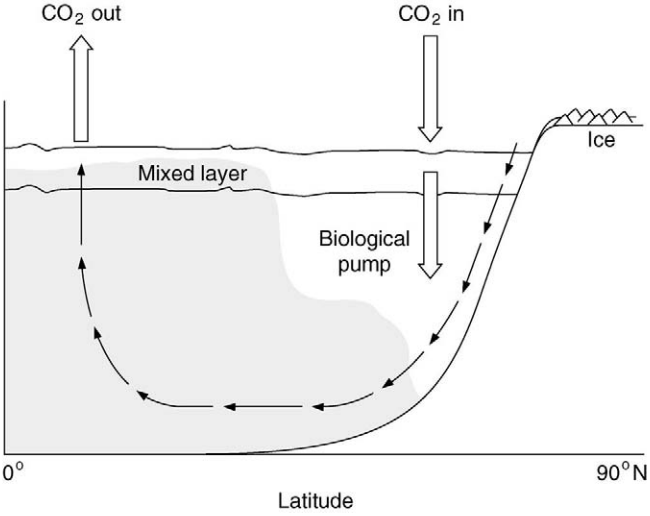


FIGURE 3 The CO_2 exchange by the ocean. The ocean is a CO_2 sink in the polar regions, aided by the biological pump. Dense water subsides, setting up a circulation pattern whereby CO_2-rich water upwells in tropical regions.

C. Uptake by Vegetation

Traditionally, ecologists have focused on measuring the biomass stocks and the net primary productivity (NPP, Table II). NPP is defined as the rate of net biological production of the phytomass, being the photosynthetic gains minus the losses of plant respiration. However, NPP is notoriously difficult to measure and is only one of the fluxes which are helpful in the analysis of the carbon cycle. The fundamental fluxes that we would like to know in addition to the NPP P_n are the net ecosystem productivity P_e and the net biome productivity P_b. They may be defined and related as follows:

$$P_n = P_g - R_a \tag{3}$$

where P_g is the gross primary productivity, corresponding to the photosynthetic rate, and R_a is the respiration of the plant component, called the autotrophic respiration:

$$P_e = P_n - R_h \tag{4}$$

The net ecosystem productivity P_e expresses the net carbon flux to the ecosystem because it allows for respiratory losses of heterotrophs R_h, including the animal and microbial consumption of the organic matter produced by plants. Micrometeorological methods, particularly eddy covariance, directly measure P_e and can be used to estimate P_n where R_h is available from measurements made in chambers attached to the soil surface.

At larger scales (geographical regions), another term is required to express the net flux after correction for disturbances and rate of change in land use which contribute to a leakage flux L:

$$P_b = P_e - L \tag{5}$$

where P_b is the net biome productivity and the leakage term L includes processes such as changes in land use, erosion, and fire.

In 1996 the International Geosphere–Biosphere Programme launched FLUXNET to promote cooperation

TABLE II

Current Estimates of Global Carbon Stocks in Vegetation and Soils and the Range of Net Primary Productivity (NPP)[a]

Biome	Area (10^6km^2)	Global carbon stocks (Gt C) Vegetation	Soils	Total	NPP (t C ha^{-1} year^{-1})
Tropical forests	17.6	212	216	428	11.0 (5.0–17.5)
Temperate forests	10.4	59	100	159	6.3 (2.0–12.5)
Boreal forests	13.7	88	471	559	4.0 (1.0–7.5)
Tropical savannas	22.5	66	264	330	4.5 (1.0–10.0)
Temperate grasslands	12.5	9	295	304	3.0 (1.0–7.5)
Deserts and semideserts	30.0	8	191	199	0.05 (0.0–0.1)
Tundra	9.5	6	121	127	0.1 (0.0–0.4)
Wetlands	3.5	15	225	240	0.9 (0.1–3.9)
Croplands	16.0	3	128	131	1.6 (0.2–3.9)
Total	135.6	466	2011	2477	

[a] From WBGU (1998).

for studies of CO_2 uptake by the terrestrial vegetation. Currently, about 70 science teams throughout the world are measuring the net uptake of carbon P_e using sensors mounted above vegetation on towers. Most of them work in undisturbed forests. Conclusions so far are as follows:

1. Most of the boreal and temperate forests in the study are sinks for carbon, in the range 0.5–8 tC ha^{-1} a^{-1}; however, most of the forests being studied are forests that have been managed for timber production and may not constitute "average" or "representative" forests.
2. The sink strength shows notable interannual variability, depending on the weather.
3. Old and undisturbed tropical forests are sinks (0.5–6 tC ha^{-1} a^{-1}).
4. In all of these forests, the net carbon gain is a small difference between two large fluxes—the incoming photosynthesis and the outgoing respiration. Of the respiratory fluxes, the flux from the soil far exceeds the flux from the aboveground plant material.

Of these conclusions, the fourth is perhaps the most vital. Examples from three forest types are shown in Fig. 4. There is a remarkable similarity in the basic pattern of fluxes in all three forest types. In all of them, the total photosynthetic uptake is almost matched by the efflux of CO_2 from the soil.

D. Deforestation and Land Use Change

The most significant change in land use is the conversion of forest to grasslands and farmlands. In former times this conversion took place in Europe and North America. In recent years it has occurred in the tropics. The area of forests in different geographical regions according to Dixon *et al.* (1994) is 4.165×10^9 ha, amounting to 359 GtC or three-fourths of the carbon stored in terrestrial vegetation (Table III). The rate of deforestation is $15–17 \times 10^6$ ha per year, most of this being in the tropics. When tropical forest is cleared to prepare the land for pasture, there is an immediate loss of carbon as carbon dioxide during combustion and smoldering. Dixon's data imply a worldwide loss of forest that corresponds to a carbon flux of 0.9 GtC per year, but most authorities now think the estimate should be higher (Fig. 1), perhaps as high as 2.0 GtC per year. In evaluating the deforestation flux it should be borne in mind that some forest disturbances including logging of large trees do not appear in satellite imagery, and this is one reason why Dixon's estimate may be too low. This type of disturbance has been called "cryptic deforestation."

Many nations have actually increased their forest area in recent years, but so far this represents only a trivial flux of carbon: 0.08 GtC per year for the United States, 0.05 for Canada, and 0.002 for Britain. In reality, the effort to manage carbon on a global scale, in order to reduce the rate of CO_2 increase, has scarcely begun.

When forests are removed to make way for pastures or arable crops, there is an effect on the climate system resulting from the change in the biophysical properties of the land surface. Forests are rougher, they reflect less radiation than farmland, and they use more water. The climatic impact of deforestation has been estimated by running global circulation models with appropriate

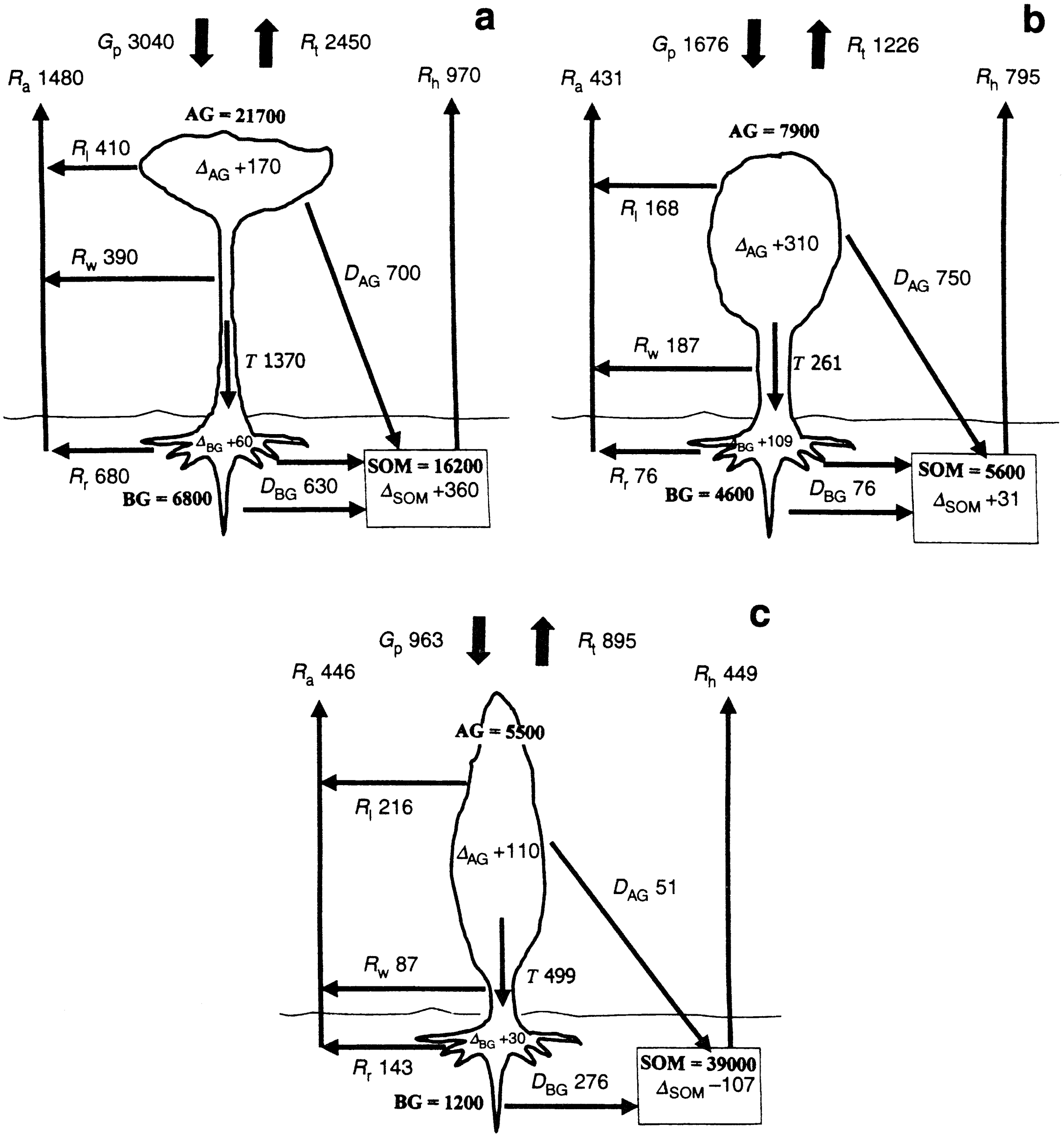


FIGURE 4 Carbon fluxes in three forest ecosystems: (a) humid tropical, (b) temperate, and (c) boreal. G_p, gross primary productivity; R_t, total respiration; R_a, autotrophic respiration; R_h, heterotrophic respiration; R_l, leaf respiration; R_w, aboveground wood respiration; R_r, root respiration; D_{AG}, aboveground detritus; D_{BG}, belowground detritus; T, belowground carbon translocation. Stocks of carbon and their annual increments are shown in bold: AG, carbon in aboveground biomass and D_{AG} is the annual increment in AG; BG, carbon in belowground biomass and D_{BG} is the annual increment; SOM, carbon in soil organic matter and D_{SOM} is its annual increment. Units are gC m^{-2} a^{-1} for fluxes and gC m^{-2} for stocks (after Malhi *et al.*, 1999).

parameterization of roughness, reflectance, and surface resistance to water loss. In tests with six different models of the Amazon basin, deforestation increased regional temperatures from 0.1 to 2.3°C, decreased precipitation from 15 to 640 mm year^{-1}, and decreased evaporation from 25 to 500 mm year^{-1}. It is too early to determine the reliability of these simulations.

E. Soils

The carbon stored in soil is the residue of thousands of years of NPP. There is uncertainty about the quantity of carbon stored as organic matter in the world's soils. Soil surveys suggest that the figure exceeds that in the vegetation by two or three times (Tables II and IV). It

TABLE III

Summary of Areas of Forests and the Carbon Stored as Biomass in the Above- and Belowground Plant Parts (not Soil Carbon)[a]

Location	Forest area (10^6 ha)	Change in forest area (10^6 ha)	Carbon stock, excluding soil (Gt C)	Change in pool (Gt C year^{-1})
High latitude				
Russia	884	−0.2	74	0.40
Canada and Alaska	488	−0.5	14	0.08
Midlatitude				
United States	241	−0.1	15	0.17
Europe	283	0.3	9	0.11
China	118	0.6	17	−0.02
Australia	396	−0.1	18	Trace
Low latitude				
Asia	310	−3.9	41	−0.70
Africa	527	−4.1	52	−0.35
Americas	918	−7.4	119	−0.60
Total	**4165**	**−15.4**	**359**	**−0.90**

[a] The period was 1987–1990 (Source: Dixon et al., 1994).

TABLE IV

Estimates of the Pools of Soil Carbon[a]

Life zone group	Area ($\times 10^{12}$ m^2)	Carbon density (kg m^{-2})	Soil carbon ($\times 10^{15}$ g)
Tropical forest, wet	4.1	19.1	78.3
Tropical forest, moist	5.3	11.4	60.4
Tropical forest, dry	2.4	9.9	23.8
Tropical forest, very dry	3.6	6.1	22.0
Temperate forest, warm	8.6	7.1	61.1
Temperate forest, cool	3.4	12.7	43.2
Boreal forest, wet	6.9	19.3	133.2
Boreal forest, moist	4.2	11.6	48.7
Tropical woodland and savanna	24.0	5.4	129.6
Temperate thorn steppe	3.9	7.6	29.6
Cool temperate steppe	9.0	13.3	119.7
Tropical desert bush	1.2	2.0	2.4
Warm desert	14.0	1.4	19.6
Cool desert	4.2	9.9	41.6
Boreal desert	2.0	10.2	20.4
Tundra	8.8	21.8	191.8
Subtotal for the study			1025.4
Cultivated land	21.2	7.9	167.5
Wetlands	2.8	72.3	202.4
Global soil carbon pool			1395.3

[a] Based on soil survey (Post *et al.*, 1982).

is noteworthy that all major biomes contain substantial carbon stocks in the soil, even humid tropical forests in which rapid decomposition rates are observed in short-term experiments.

Soil organic matter can persist for a long time. Litter fall, exudates, leachates, and dead roots enter the soil and provide the substrate for the soil microflora. Decomposition is a multistage process, involving a succession of insects, bacteria, and fungi, operating on the various biochemical constituents of the detritus. The products of decomposition are CO_2, H_2O, and mineral ions. The latter are usually captured by the roots and mycorrhizas. The CO_2 accumulates in the pores of the soil and diffuses to the atmosphere as the heterotrophic respiration R_h, forming perhaps half of the total CO_2 flux from the soil. The remainder originates from root respiration.

As the dead remains of plants and animals (known as detritus or simply litter) are broken down they assume a different chemical and physical form, and when the cellular structure is no longer discernible the material is termed humus. Humus is chemically heterogeneous and contains some fractions which are extremely resistant to decay, particularly those which originate from lignin. The acid-soluble fraction of humus is termed fulvic acid, and the insoluble fraction is termed humic acid. These fractions are chemically heterogeneous. They are transported vertically in the soil by rain and they contribute to the long-lived soil organic matter termed the "recalcitrant fraction." In storms they may be transported laterally to rivers, to which they impart a brown coloration, and they may finally enter the sea. Schlesinger presented an example of the stocks and pools in a grassland, showing typical pool sizes and resident times applicable to the various fractions (Fig. 5).

Current interest centers on the effect of temperature on the decomposition rate because the carbon balance of the terrestrial surface is very sensitive to increases in respiration, and it is expected that climate warming will increase the respiration rate. There are reports of ecosystems "flipping" from being sinks to being sources, and there is one report of an unexpectedly large respiration flux derived from "old" carbon in forest soils.

In laboratory studies the relationship between respiration and temperature can generally be approximated

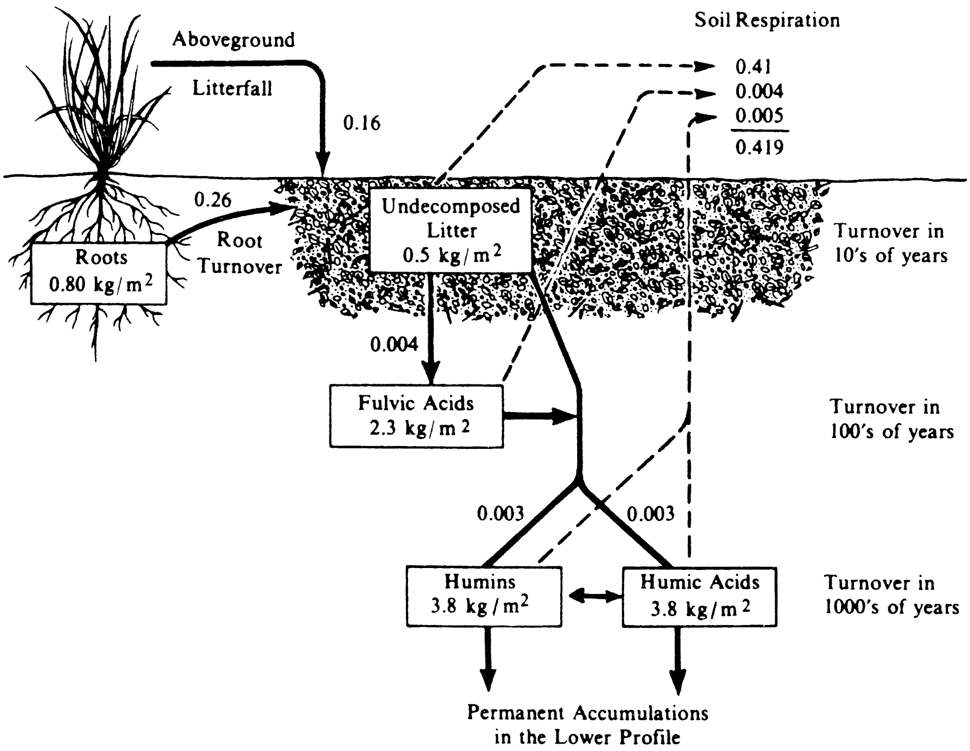


FIGURE 5 Annual carbon flows in the 0- to 20-cm layer of a grassland soil. Carbon stocks are in units kg C m^{-2}, and fluxes are in kg C m^{-2} a^{-1}. The soil contains 10.4 kg C m^{-2} in the 0- to 20-cm layer (after Schlesinger, 1997).

by an exponential function, with a Q_{10} of approximately 2. However, much higher Q_{10} values, ranging from 2 to 8, have been reported from chamber measurements of CO_2 efflux from the soil. This matter is currently one of great research interest because approximate calculations suggest that the soil efflux of carbon is likely to exceed the photosynthetic influx in a warmer world, thus causing complete loss of the terrestrial sink.

One small fraction of the soil carbon is elemental carbon derived from fire, occurring as fine deposits and as charcoal. This fraction is not decomposed by fungi and bacteria. In a regime of burning, elemental carbon can be expected to accumulate over centuries.

F. Rivers

The surface area of river and inland waters is not very great compared with that of land and ocean, and therefore the direct exchanges of carbon between rivers and atmosphere are not especially important. However, the rivers of the world are conduits between the land and ocean, carrying large amounts of carbon and nutrients to the sea. Estimates based on analyses of the organic material entering the sea from the world's rivers have been made based on survey data from a large proportion of the main rivers (Table V).

Particular inorganic carbon (PIC) originates mainly from the mechanical erosion of sedimentary rocks. Dissolved organic carbon (DOC) comes from soil leaching. Exudates from roots, or leaching from leaf canopies, releases simple organic compounds such as organic acids, sugars, and alcohols, which decompose rapidly (hours or days). However, humic materials have a long lifetime (hundreds and thousands of years), and are often present as colloids. In analysis, dissolved and particulate fractions are somewhat arbitrarily separated by a 0.5-mm filter. The captured part is called particulate organic carbon (POC). POC includes material derived from soil and peat by erosion, but there are also coarse fractions as the debris of plants living in the riparian zone.

From the viewpoint of the carbon cycle, the flux of carbon derived from the atmosphere (rather than geologically derived) is the relevant quantity. It is given as the sum of the atmospheric component of dissolved inorganic carbon (DIC) + DOC + atmospheric POC, and amounts to 0.48 GtC year^{-1} (Table V). It is less than 1% of the global NPP.

G. Geological Processes

There is an immense quantity of carbon stored as carbonate rocks—perhaps as much as 65×10^6 Gt. The chemical forms predominating are $CaCO_3$, occurring in limestone in the mineral forms of calcite and aragonite; $MgCO_3$ (magnesite); $CaMgCO_3$ (dolomite); $MnCO_3$ (rhodochristite); and ironstone or siderite ($FeCO_3$). Over geological timescales, carbonates have been precipitated both biotically and abiotically. Corals, gastropods, and clams have shells made of aragonite; foraminifera, and coccolithophores; some gastropods and clams

TABLE V

Transfer of Carbon from Rivers to the Ocean[a]

	Source	Age of carbon (years)	Flux (Gt C a^{-1})	A	B	C	D	E	F
PIC	Geological	10^4-10^8	0.17	*					
DIC	Geological	$10^4–10^8$	0.14		*	*			
	Atmospheric	$0–10^2$	0.245		*	*			
DOC	Soils	$10^0–10^3$	0.20			*			
	Pollution	$10^{-1}–10^2$	0.01 (?)					*	
CO_2	Atmospheric	0	0.02–0.08		*	*	*		
POC	Soil	$10^0–10^3$	0.10	*					*
	Algal	10^{-2}	0.01				*		*
	Pollution	$10^{-2}–10^0$	0.015					*	
	Geological	$10^4–10^8$	0.08	*					*

[a] PIC, particulate inorganic carbon; DIC, dissolved inorganic carbon; DOC, dissolved organic carbon; POC, particulate organic carbon. Sensitivity to change is indicated by letters A–F as follows: A, land erosion; B, chemical weathering; C, global warming and ultraviolet radiation; D, eutrophication; E, organic, pollution; and F, basin management (From Maybeck, 1993).

have shells made of calcite; and starfish have exoskeletons made of high-magnesium calcite.

Over long periods of time, carbonates and silicates in the marine sediment are subducted in the earth's crust, where they are metamorphosed. In this process, primary minerals are created and CO_2 is released and returned to the atmosphere from volcanoes and fumeroles, currently at a very low rate (0.02–0.04 Gt a^{-1}) compared to that of fossil fuel emissions. This is sometimes known as the geologic carbon cycle. It takes 100–200 million years for the ocean crust to circulate.

The composition of volcanic gases is highly variable. For example, Kilauea in Hawaii was 2.3% by volume CO_2, Stromboli in Italy was 30%, and the eruptions from Lake Nyos in Cameroon may have been almost pure CO_2. Likewise, the total quantities of carbon from active volcanoes are very variable: Mt. St. Helens was 3 ktC day^{-1}, Etna was 30 ktC day^{-1}, and Nyiragongo was 45 ktC day^{-1}.

The weathering of rocks consumes CO_2. In the past, when CO_2 concentrations were higher, exposed rocks were weathered. Currently, the weathering process is more intense in the soil. There, CO_2 concentrations are much higher than the 0.036% found in the atmosphere, typically reaching several percent. This is the result of respiration of roots and microbes. Thus, carbonic acid H_2CO_3 and its dissociation products bicarbonate, carbonate, and hydrogen ions are formed (Eq. 2).

Carbonic acid weathers minerals in several ways, as illustrated in the following examples. In the first case, Na-feldspar is weathered and the carbon emerges as a product in the form of bicarbonate. Another product is the clay mineral kaolinite $Al_2Si_2O_5(OH)_4$:

$$\begin{aligned} 2NaAlSi_3O_8 + 2H_2CO_3 + 9H_2O \Rightarrow{} & 2Na^{++} \\ & + 2HCO_3^- + 4H_4SiO_4 + Al_2Si_2O_5(OH)_4 \end{aligned} \tag{6}$$

In the second example, calcium carbonate is weathered, and once more bicarbonate is produced:

$$CaCO_3 + H_2CO_3 \Rightarrow Ca^{++} + 2HCO_3^- \tag{7}$$

Bicarbonate produced in this way may enter streams and rivers, finally contributing to the DIC fraction of the carbon in the ocean.

H. Fossil Fuel Reserves

The total extractable fossil fuel reserves have been estimated as 4000 GtC, more than 60 times the amount of carbon in the present-day biomass and enough to last for several hundred years of fossil fuel burning under present-day rates. Fossil fuels comprise coal (3500 GtC), oil (230 GtC), and gas (140 GtC). Coals are formed from partially decomposed plant material, the product of photosynthesis in (mainly) the period from the Carboniferous to the Cretaceous (65–355 million years ago). In the process of coalification, the plant material is concentrated so that the carbon content increases from 50 to 95%. Oil is derived from pelagic material and contained in shales and limestones, e.g., the shales of Iraq–Iran oilfields, the Jurassic limestones of Saudi Arabia, and the Cretaceous limestones in Venezuela. Both the oil source rock and coal-bearing strata may produce associated gas. Crude oil is a mixture of hydrocarbons of the paraffin series, which are straight chains with the formula C_nH_{2n+2}, and the naphthene series, which are carbon ring compounds C_nH_{2n}. The first four members of the paraffin series are gaseous and together constitute natural gas: methane, ethane, propane, and *n*-butane.

Peat is a precursor of fossil fuel. Peat occurs in boggy regions and is harvested for fuel and for horticultural use. Peatlands may contain as much as 250 GtC.

The rate of use of fossil fuel has increased sharply in recent years. Although the per capita emission rate is far greater in rich countries, these countries kept their overall emission rates more or less constant in the period 1990–1996, whereas the poorer, less industrialized countries increased their rate (Table VI). Some poorer countries, notably China, have enormous fossil fuel reserves and a low per capita emission rate. Currently, it is too early to say whether the Kyoto Protocol, which attempts to limit fossil fuel emission rates country by country, will be effective in controlling the overall rate of global emissions (see Section V).

II. RESPONSE OF THE OCEANIC SINK TO CLIMATE CHANGE

The ocean contains a large volume of water (1.3×10^9 km^3), with a correspondingly large capacity to dissolve CO_2. Given the high solubility of CO_2 in water, we would not expect the ocean to become saturated with CO_2 until it had accumulated in excess of 10^5 Gt carbon, much more than it has today. However, exchange of gas between the atmosphere and the ocean is a slow process because only the surface layer, a few meters deep, is thoroughly mixed by wind action, and there is very poor mixing with the bulk of the ocean. This poor mixing is revealed from studies of the radionu-

TABLE VI

Fossil CO_2 Emissions Breakdown[a]

Year	Annex 1 (GtC per year)[b]	Non-Annex 1 (GtC per year)	Bunker fuels (GtC per year)[c]	Total (GtC per year)
1990	3.91	2.09	0.11	6.11
1991	3.79	2.27	0.12	6.18
1992	3.71	2.24	0.14	6.08
1993	3.63	2.29	0.13	6.05
1994	3.68	2.48	0.13	6.29
1995	3.80	2.57	0.14	6.51
1996	3.87	2.66	0.14	6.67

[a] Fossil CO_2 emissions given are from fossil fuel combustion (including gas flaring), cement production, and other nonfuel carbon sources.

[b] Annex 1 countries are Australia, Austria, Belarus, Bulgaria, Canada, Czechoslovakia, Denmark, European Economic Community, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Latvia, Lithuania, Luxembourg, The Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Russian Federation, Spain, Sweden, Switzerland, Turkey, Ukraine, United Kingdom, and United States.

[c] Bunker fuels are those used in air and marine international transport.

clides released from bomb tests in the 1960s. Calculations based on the circulation rate suggest that it would take about 1000 years for equilibrium between the atmosphere and ocean to occur.

The response of the ocean sink to climatic change is not well established. It is presumed that in preindustrial times there was more or less a steady state, i.e., sink strength was zero. As the CO_2 concentration steadily increased from about 280 ppm in 1800 to present-day values in excess of 360 ppm, CO_2 began to be taken up by the ocean; however, despite the huge capacity available, the rate of uptake has been limited by the diffusional process at the sea surface so that there is essentially a lag in the response. Thus, if fossil fuel emissions were to abruptly cease, the sink would continue to operate for a long time (a century perhaps) because the concentration gradient driving the process would persist. This dynamic feature of the oceanic sink tends to dominate the more obvious short-term responses to temperature and CO_2 concentration.

The solubility of gases in water is a strong function of temperature, and it should be kept in mind in discussions of global change that all gases are considerably less soluble in water as the temperature increases (Table VII). Thus, a 1°C increase in temperature over the range 10–15°C is expected to reduce the sink strength by about 3%. A doubling of CO_2 concentration is expected to increase the diffusion gradient from atmosphere to sea, and hence the sink strength, by much more. Fluctuations in sea surface temperatures that result from climatic phenomena such as El Niño can be more than 1°C, but El Niño reduces the rate of upwelling of the Pacific Ocean, causing less efflux than in a normal year, and this overwhelms the expected effect of temperature. There are several other potentially critical processes. For example, increasing temperature will increase the melting rate of ice in the polar regions and thus increase the thermohaline cycle and the rate of upwelling in tropical waters. This effect is not well understood, is difficult to model, and is much more important than the direct effect of average temperature on the dissolution of CO_2. It is becoming increasingly important to understand the marine biological processes in the context of

TABLE VII

Solubilities of Gases in Water at Atmospheric Pressure (101 kN m^{-2})[a]

Gas	Temperature (°C)			
	10	15	20	30
Carbon dioxide	1.16	0.99	0.89	0.65
Nitrogen	0.018	0.016	0.015	0.013
Oxygen	0.037	0.033	0.030	0.026
Carbon monoxide	0.028	0.025	0.023	0.020

[a] Units: m^3 of gas per m^3 of water at the stated temperature.

the “sinking” of carbon and the operation of the biological pump. The pump is unlikely to be directly stimulated by CO_2 because the phytoplankton is believed to be rate limited by nutrients rather than by carbon. A factor which might stimulate the biological pump is the additional deposition of materials to the ocean. Approximate calculations of this effect have been made based on the Redfield ratio.

The Redfield ratio is derived from the chemical analysis of phytoplankton and expresses the atomic ratios of biomass. There tends to be a constant ratio of C : N : P such that marine biomass can be written as $C_{106}N_{16}P$. Because upwelling waters contain far more carbon relative to N and P than is required in the Redfield ratio, it is supposed that C concentration is not limiting. Thus, we may predict how much extra biomass might possibly be synthesized if the supply of rate-limiting N or P were enhanced, either by additional upwelling or by anthropogenic sources. It is well established that the deposition of nitrogen as nitrate and ammonium from anthropogenic sources has increased, particularly during the past few decades, thus augmenting the nitrate that was always present in the atmosphere as windblown dust derived from the land. The anthropogenic rates of dry and wet deposition are especially enhanced over the land of Europe and eastern North America, but atmospheric dispersal will have spread the emissions even to the ocean. It has been estimated that the “excess” nitrogen from anthropogenic sources could have augmented the nitrogen supply to the ocean by 17×10^{12} gN a^{-1}, which is the Redfield equivalent to an additional 0.1×10^{15} gC a^{-1}. Thus, the stimulation of the biological pump by anthropogenic nitrogen is currently only slight (see Fig. 1).

Of course, nitrogen is abundant in the atmosphere as N_2, a gas which is sparingly soluble in water (0.018 m^3 nitrogen in 1 m^3 water at 10°C). In fresh water, nitrogen fixation is achieved by aquatic nitrogen-fixing organisms, particularly the cyanobacteria. In seawater, nitrogen fixation proceeds at a slow rate as a result of deficiencies of micronutrients (notably Mo and Fe). A most interesting recent discovery has been that phytoplankton in the Pacific Ocean are stimulated to assimilate additional CO_2 by the experimental addition of Fe.

In concluding this section, the reader should realize that our understanding of the processes which underlie the working of the ocean sink is incomplete. Some comparisons with the state of the art in terrestrial ecosystems can be made. Although terrestrial ecosystems are very heterogeneous, they are amenable to mapping and detailed description, and there is a wealth of knowledge about carbon fluxes within them. On the land it has become possible to directly measure the CO_2 fluxes over entire ecosystems, whereas marine scientists have to use indirect methods. For example, the oceanic gas exchange at small temporal and spatial scales is usually measured in a glass bottle which is lowered from a ship. At large scales, the fluxes have to be calculated from CO_2 concentrations and wind speed-dependent exchange coefficients. A very large effort is required to survey the entire ocean, as has been done during the International Geosphere–Biosphere Programme, to derive global maps of concentration and calculated sink strength, and there is inevitably a large reliance on models. Progress in evaluating the current sink strength of the ocean has largely been made by atmospheric scientists, but predicting future sink strength requires an understanding of the biological pump—an understanding that marine biologists must provide.

III. RESPONSE OF THE TERRESTRIAL SINK TO CLIMATIC CHANGE

Insight into the carbon-accumulating capacity of ecosystems comes from studies of succession. For example, on glacial moraines in Alaska vegetation and soil develop over a period of several decades. In this period, discrete stages of vegetation occur as different life forms colonize the land. Initially, only lichens and mosses are able to withstand the resource-poor environment, but later grasses, herbs, and trees colonize. Some of the organisms have special attributes; for example, the lichens are capable of fixing atmospheric nitrogen, and the trees give rise to a spatial heterogeneity in microclimate that encourages biodiversity. Over 100 years the ecosystem aggrades, and the soil alone accumulates 0.4 tC ha^{-1} a^{-1} and about 0.02 tons of nitrogen ha^{-1} a^{-1}. Thereafter, the rate of change is slow, and it is common to assume that a steady “mature” state is reached at which photosynthetic gains and respiratory losses are equal. Elsewhere, when agricultural land is abandoned, we observe a regrowth of vegetation—first as scrub then as forest. Examples from the tropics suggest an accumulation rate in the biomass of several tons of carbon ha^{-1} a^{-1}, with a tendency to reach a ceiling after about 100 years. Unfortunately, no one has properly observed this ceiling, and it is possible that a steady state does not occur. In peatlands carbon can continue to accumulate for thousands of years until a profound disturbance occurs. This calls into question the normal assumption that vegetation reaches a steady state when it is mature.

In managed ecosystems, the natural tendency of ecosystems to accumulate carbon in the juvenile stage is exploited by farmers and foresters, who intervene by picking, cutting, chopping, or burning. This intervention creates much disturbance to the soil, and in some cases losses of carbon to the atmosphere may be significant as a result. However, over entire crop cycles there may well be a net accumulation of carbon in the soil as humus and, in the case of forests, an accumulation not only in the soil but also on the land as buildings, furnishings, and other artifacts.

A. CO_2 Effects on Photosynthesis and Growth

The response of photosynthesis to CO_2 and temperature has been much studied in recent decades. Commercial growers have long known that elevated CO_2 increases productivity under green house conditions, and good measurements of photosynthesis of leaves, made in short-term experiments as early as 1960 by Gaastra, showed a strong positive effect of elevated CO_2. Those who work on global modeling of the impact of elevated CO_2 on plants like to use a single parameter β to describe the extent of the stimulation of NPP brought about by doubling the CO_2 concentration, but plant ecophysiologists see difficulties in this simplistic approach because it ignores many important factors. Hundreds of species have been grown experimentally at twice-normal CO_2 concentrations, but firm conclusions about the value of β have been elusive. There are several reasons for this. First, species differ in their response to elevated CO_2, and for any single species there are changes in sensitivity over the life cycle as acclimation occurs. There are also likely to be variations according to the nutrient and water supply, and there may be complicating effects when several species are growing together and thus competing with each other.

Experiments on the impact of elevated CO_2 on plants are conducted using several experimental techniques. Controlled environment rooms and outdoor chambers cover an experimental area of a few square meters and usually permit plants to be grown only for quite limited periods (weeks to a few years) at elevated CO_2. Plants may be rooted in natural soil, but more often they have been planted in containers which have a restrictive effect on the supply of nutrients and water. There are exceptions, in which closed chambers for ecosystem studies have operated over several years, especially in tundra and grassland ecosystems. Sometimes, very large, greenhouse-sized chambers have been used, as in the case of the Biosphere-2 experiment in Arizona. However, all chambers produce unwanted microclimatological effects, and to overcome these artificial influences plants have been exposed to high CO_2 in completely open conditions in the field, in their normal cultivation conditions (the so-called FACE approach—free air carbon dioxide experiment). Recently, FACE experiments have been established at forest sites, and it is expected that these will yield the most valuable information about the response of major ecosystems to climate change, although they consume large amounts of gas and are thus costly to run. Finally, it has sometimes been possible to locate suitable natural sources of CO_2, as fumeroles or CO_2 springs, where mature vegetation is exposed permanently to elevated CO_2. There are also drawbacks to this approach because the concentration is uncontrolled, and there may be traces of toxic gases.

A few conclusions may be drawn from the very large and often conflicting literature:

1. In trees, the stimulatory effect of twice-normal CO_2 on growth was reported in one review to be 38% in coniferous trees and 63% in broad-leaved species. However, the stimulation declines as the tree ages. This effect of age has been shown in sour-orange, birch, and Mediterranean oaks growing near high-CO_2 springs in Italy.
2. In herbaceous crops, the stimulatory effect was reported to be 16% on average, with some species being stimulated by as much as 80%, whereas others showed inhibition of growth.
3. There is often a downregulation of photosynthesis when plants are grown at elevated CO_2. This takes several forms. The most obvious is a decline in the activity of the photosynthetic enzyme Rubisco (ribulose bisphosphate carboxylase oxygenase). There may be a reduction in the number of stomata per area of leaf. Thus, a doubling of CO_2 concentration does not usually produce a doubling of the photosynthetic rate: More often the stimulation is in the range 20–60%.
4. Most plants show an increase in water use efficiency (WUE) at elevated CO_2. WUE is defined as the carbon taken up divided by the water used.

It is not clear how these conclusions are to be scaled up to whole ecosystems. Models play an important role in the use of experimental data to make predictions at large scales, although much of the earlier data are not useful in parameterizing the models because plants were grown in wholly artificial conditions.

IV. HISTORY OF THE CARBON CYCLE

Carbon and other elements are produced in stars by nuclear fusion. Stars explode, and under the attractive force of gravity the debris from an exploded star (supernova) condenses to make a sun and planets. In our solar system this is believed to have happened 4.6 billion years ago, from the radiodated age of the oldest rocks found on the moon and as meteorites. Over a relatively short period, perhaps 100 million years, additional mass would have been accumulated from planetesimals and meteors, including meteors known as carbonaceous chondrites, which contain an abundance of the lighter elements including carbon. The carbon content of these is 0.5–3.6% and the nitrogen content is 0.01–0.28%. The planet would be hot inside, as a result of pressure and also as a result of radioactive decay of the larger atoms from the debris of the supernova. There would be considerably sorting and reprocessing of the elemental material to form minerals. An early planet would have no atmosphere, but outgassing from the hot interior would result in a mixture of gases containing water, nitrogen, carbon dioxide, hydrochloric acid, and sulfur dioxide—a mixture probably not unlike modern volcanic gas mixtures. Cooling would result in the water vapor condensing to form the ocean, although this could have been lost several times by meteorite bombardment which is believed to have been very intense until about 3.8 billion years ago. Gases such as SO_2 and HCl are extremely water soluble and form acids in aqueous solution. The acidity would have been neutralized by reaction with minerals at the surface. Nitrogen gas, N_2, is not very soluble in water and would have thus formed the bulk of the atmosphere. CO_2 is quite soluble in water (Table II) and would have eventually formed an equilibrium concentration.

The oldest fossils known to man are stromatolites, which are found as laminated calcareous deposits and are up to 3.5 billion years old. They are made up largely of filamentous organisms that, by comparison with modern analogs, are believed to be mainly the photosynthetic prokaryotes cyanobacteria. The early communities may have resembled those which are now living at high temperatures in hydrothermal vents or “algal mats” in coastal regions. Photosynthesis may be even somewhat older than 3.5 Ga because the carbon isotopic signature of organic matter as old as 3.8 Ga shows ^{13}C to be depleted relative to its abundance in dissolved bicarbonate, a result of the greater affinity of the carbon-fixing enzyme Rubisco for $^{12}CO_2$. Two analogous sorts of photosynthesis may have occurred in ancient seas: first, the reaction performed today by the sulfur bacteria, in which pure sulfur is deposited:

$$CO_2 + 2H_2S + \text{sunlight} \rightarrow CH_2O + 2S + H_2O. \quad (8)$$

Second, photosynthesis produced today by green plants, essentially the photochemical splitting of water to release oxygen:

$$2H_2O + CO_2 + \text{sunlight} \rightarrow CH_2O + O_2 + H_2O \quad (9)$$

Oxygen accumulates only where the organic matter becomes buried and thus protected from decomposition or fire. It is presumed that the total oxygen that has ever been released is matched atom for atom by the storage of reduced carbon (coal, oil, and hydrocarbon gases), and it is approximately true that if all the known organic matter were to be burned the oxygen concentration would decrease to zero. In reality, biomass does not burn when the oxygen concentration decreases below about 15%. Lovelock noted that the current concentration of oxygen on Earth (21%) may be subject to homeostasis. Any more than 21% O_2 and biomass would burn uncontrollably, thus oxidizing more carbon and bringing the oxygen concentration back to the initial level. At much less than 21% O_2, fires do not burn, and thus oxygen can accumulate.

The oxygen took a long time to accumulate in the atmosphere as it reacted immediately with exposed minerals, especially iron, to produce Fe_2O_3. Banded deposits of Fe_2O_3 occur in rocks that are 2.5–3.0 Ga old. It was not until about 2 Ga ago that an oxygen-rich atmosphere developed, enabling aerobic heterotrophic life to evolve and forming a ultraviolet (UV)-shielding layer of ozone in the stratosphere. Low UV is an important condition for the colonization of the land.

So far, I have outlined the modern theory for the establishment of the carbon cycle and the generation of the oxygen-rich atmosphere, which distinguishes Earth from its sister planets Mars and Venus. These planets are lifeless, they have lost their oceans, and they contain an atmosphere which is almost pure CO_2.

Subsequent evolution of the carbon cycle, over a period of 2 billion years, is not well-known. A speculative reconstruction of partial pressures of CO_2 and O_2 has been made based on the oxidation state of iron and manganese in sedimentary and metamorphic rock and observations on the body size of certain groups of animals (Fig. 6). Other authors have examined the sedimentary record of organic carbon to calculate the rate of accumulation of carbon stocks during the past 0.6 Ga. It increased from a low value at 0.5 Ga to a peak

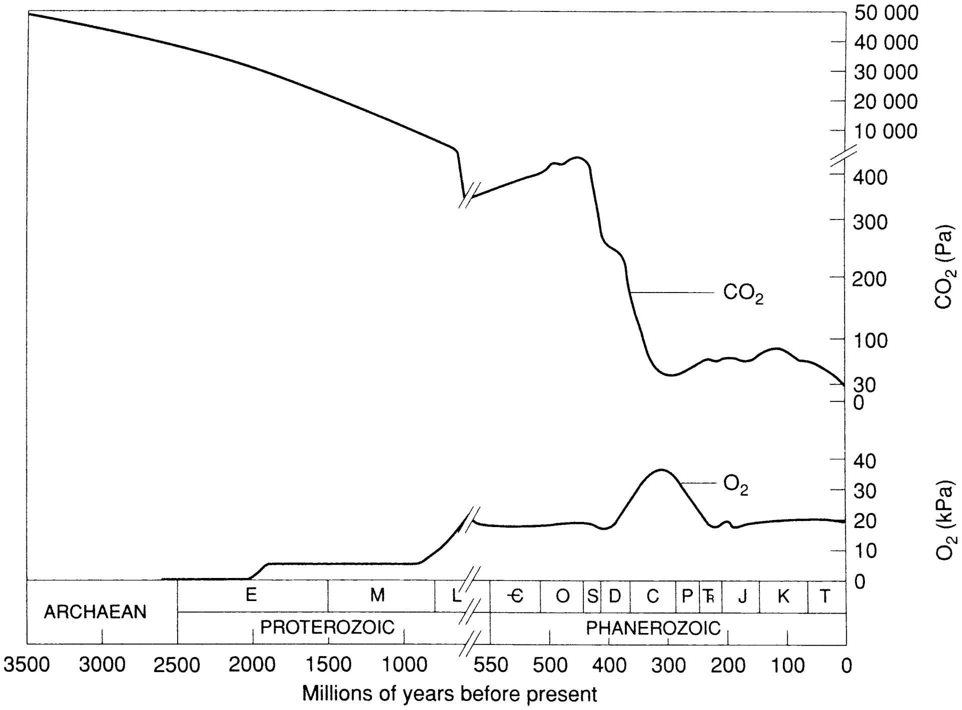


FIGURE 6 The estimated variations in partial pressures of CO_2 and O_2 over 3.5 billion years of Earth history (after Berner, 1993; Falkowski and Raven, 1997).

in the carboniferous. This is a time when the land flora was well developed over much of the earth, based on spore-bearing plants. Fish and mollusks were well established, and amphibians lived in tropical swamps. Reptiles were beginning to evolve.

As we consider recent times, it is clear that the carbon cycle has not been in a steady state over periods of tens of thousands of years. Fluxes between atmosphere and vegetation, and atmosphere and ocean, have occurred in response to changing temperature and changing vegetation cover. Evidence for the changing nature of the cycle in relation to temperature and vegetational cover comes from cores extracted from ice or firn in polar regions, especially in Greenland and Antarctica. The cores contain bubbles of trapped air, which are released when the core is thawed. Concentrations and isotopic ratios of CO_2 and CH_4 can be measured, and temperature can be estimated from the isotopic signature of oxygen in water. This type of record has shown clearly the correlative link between CO_2, CH_4, and temperature.

The most recent record, from Taylor Dome in Antarctica, is especially free from known artifacts and covers the postglacial period (the past 11,000 years). Following the retreat of the ice, the CO_2 concentration decreased slightly from 265 to 260 ppm. Thereafter, it increased to about 280 ppm before the period of industrialization and extensive deforestation. This pattern is mirrored by the isotopic signature of the carbon in CO_2, which at first increases and then decreases (Fig. 7). These changes in isotopic signature can only be achieved through massive changes in gas exchange with the land associated with terrestrial photosynthesis because exchanges with the ocean are not accompanied by isotopic discrimination. The changes from 11 to 7 ka BP are best interpreted as a consequence of the rapid growth and spread of vegetation when the temperature increased and the ice retreated. The climatic optimum for vegetation in the Northern Hemisphere, documented in the pollen record, occurred at about 7 ka BP. Thereafter, conditions became colder and drier; in Africa, for example, there was a change from grassland and savanna to desert. This seems to have been accompanied by a transfer of carbon from land to atmosphere to increase the CO_2. There may have also been other changes. For example, the increase in sea surface temperature would have caused a flux of carbon from the ocean to the atmosphere, but such changes are not accompanied by any fractionation because only terrestrial C_3 photosynthesis is capable of substantial fraction-

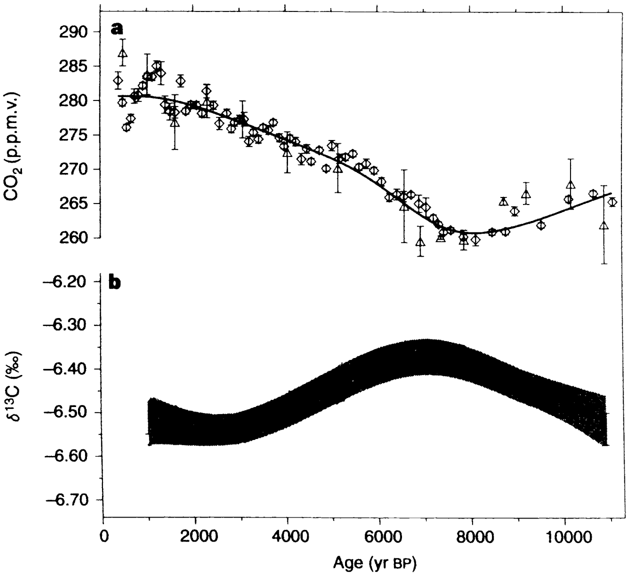


FIGURE 7 CO$_2$ concentrations and stable isotope ratios found from bubbles in an ice core from Taylor Dome in Antarctica, representing atmospheric concentrations of CO$_2$ over 10,000 years and demonstrating the contribution of terrestrial photosynthesis to changes in concentrations (reproduced with permission from Indermühle *et al.*, 1999).

ation. It is possible to conclude that changes in fluxes to and from the ocean were of minor importance during the period.

Rates of change of CO$_2$ concentration in the atmosphere, revealed by the ice core record, were always much slower than the current growth rate of 3.1 GtC a^{-1}.

V. MANAGING THE CARBON CYCLE

Responding to concerns that human activities are increasing atmospheric concentrations of greenhouse gases such as carbon dioxide and methane to levels which threaten to cause damaging global warming, most nations of the world signed the United Nations Framework Convention on Climate Change in 1992. They pledged to work toward stabilization of emissions. Following subsequent negotiations, the Kyoto Protocol (in late 1997) committed nations to legally binding reductions in emissions of six greenhouse gases: carbon dioxide, methane, nitrous oxide, hydrofluorocarbons, perfluorocarbons, and sulfur hexafluoride. These reductions will be counted from 2000 to the "commitment period" between 2008 and 2012. To date, 174 countries have ratified the treaty, and industrial countries are legally bound to decrease emissions by 5.2%. The European Union has agreed to reduce emissions 8% and the United States has agreed to 7%. Less industrialized countries have had difficulties in agreeing to any reductions. China, for example, has stated that it will not reduce emissions until it achieves the level of a medium-developed country. One notable industrialized country (Japan) is likely to achieve its target of 6% by building more nuclear reactors.

Under the terms of the treaty, reductions by the industrialized countries (called Annex 1 countries) may be met either by making a real reduction in emissions or by increasing the sequestration of some terrestrial ecosystems.

There are good prospects of reducing emissions in the short term by improvements in energy conservation and in the longer term by devising an energy strategy which is less dependent on fossil fuel. Since the energy crisis of the early 1970s, the developed world has been slow to embrace new energy technologies because there are still large supplies of fossil fuel. Even in Europe transportation is still dominated by the CO$_2$-emitting automobile rather than by public transport which emits less carbon per capita per kilometer traveled (Table VIII). The United States has a human population of 270 million, which is almost matched by a vehicle population of 210 million. The annual average mileage is 11,800; total gasoline consumption is 146 × 10^9 gallons, and the carbon emissions exceed 1 tC a^{-1} per head of population.

Transport, however, is often used as a scapegoat. It is the most visible form of CO$_2$ emission but not the largest. In the United Kingdom, the per capita emissions expected in the Year 2010 are as follows (tC a^{-1}): industry and business, 1.45; energy supply, 0.98; domestic, 0.78; transport, 0.65; agriculture and forestry, 0.43; and public sector, 0.16.

TABLE VIII

Approximate Carbon Costs of Traveling

Mode	Carbon cost (kg C/persons/km)	Comment
Automobile	0.059	Occupancy assumed to be 1.8 people per car; UK conditions
Bus	0.024	In London
Subway/metro	0.030	In London
Train	0.009	In United Kingdom
Plane	0.067	Short haul
Plane	0.044	Long haul

A. Forest Sequestration

The rate at which CO_2 is increasing in the atmosphere could be moderated by (i) slowing down deforestation and protecting existing forests, (ii) expanding the forested area, and (iii) substituting fossil fuel-based products with those made of biomass.

The prospect of growing trees to sequester carbon is appealing for many reasons. First, the technology is established and "safe." Second, the "Kyoto" forests thus created may have significant other benefits, such as environmental enhancement, increase in biodiversity, and ultimate economic reward when timber is harvested. Forests are potentially useful sinks for carbon but vary greatly in their capacity to absorb carbon dioxide. Well-managed plantations show an uptake of 1–5 tC ha^{-1} per year over their life span. A simple calculation, however, will show that the forests would need to be on a large scale to be globally useful. If we wanted to absorb the total annual emissions in excess of 6 Gt, and if we assume fast-growing plantations (5 tC ha^{-1} per year), we would need more than 10^9 ha of land, which is 40% more than the area of the entire Amazon basin. This argument shows that we must regard forests as only one of several measures that must be taken to reduce CO_2 concentrations in the atmosphere. Third, there is the question of what should be done with the trees at the time of harvest (50–150 years in the future). Some forest products (newspaper and chopsticks) are short-lived, but many others are long-lived (buildings, furniture, and carvings). A shift from the former to the latter is required. Timber could replace cement and concrete to save the CO_2 emissions that occur from the cement industry.

Currently, the Kyoto Protocol allows only certain types of forest. In order to qualify, a forest must be within an Annex 1 country and must be a new forest, constituting "afforestation, reforestation, and deforestation" since 1990. The accounting will be done in the commitment period defined as 2008–2012. Thus, if a country were to establish a large forest in 2000, and if the forest had grown and accumulated carbon by 2012, it would receive "forest credits" which could be set against its emissions. Under the terms of the treaty, these credits could be traded, in much the same way as stocks and shares are traded.

How would governments control emissions from industry? Currently, this is not very clear. There could be taxes on carbon emissions or (more likely) the relevant government agency would issue pollution permits. These permits would define the upper limit of greenhouse gas emissions, beyond which the industry would be in breach of the law. An industry with less than its limit would accumulate a carbon credit, which could be traded with another industry that exceeded its limit. Industries could acquire land for afforestation and thus receive forest credits.

The mode of operation of such schemes is still a matter for discussion. The Intergovernmental Panel on Climate Change is expected to report on the matter in 2000. The following general difficulties and specific loopholes in the operation of the protocol have been noted:

1. Many environmentalists believe that forest sequestration avoids the fundamental issues of how to reduce emissions and how to develop sustainable world economies. There is an important question of whether substantial reductions in emissions are possible without sacrificing economic competitiveness and growth. In the United States, for example, which has a reduction commitment of 7%, members of Congress who opposed the treaty were quick to draw attention to the likely economic impacts that might occur if the United States were to attempt to meet its emission reduction commitments.
2. The Kyoto forests need to be closely monitored, perhaps by satellite surveillance but also by site visits, to verify that carbon has indeed been sequestered as claimed. This requires measurements of belowground as well as aboveground carbon stocks. Currently, scientific knowledge of carbon accumulation over the whole lifetime of a forest stand, including planting, tending, and harvesting, is probably not adequate. What government agencies would be responsible for monitoring, and what would be the economic cost of monitoring?
3. There are loopholes in the protocol that permit counterproductive actions to occur. The discontinuity of commitment periods could be exploited by felling the older forests between periods, thus enabling the accounting period to start again. Moreover, forests that have been disqualified because they were planted in earlier times might now be prematurely felled so that planting can occur within the accounting period.
4. Another way to manage the carbon cycle is to reduce tropical deforestation, but this is currently not allowed for in the protocol. Deforestation is likely to continue at current rates and even accelerate. Would it not be better to explore the management of the carbon cycle by prevention of deforestation in the tropics? In fact, there are several such initiatives in the tropics that predate Kyoto. For example, power utilities, mining companies, automobile companies, and other commercial enterprises as well as the World Bank are paying

farmers or government agencies in Bolivia, Mexico, Brazil, and Uganda for carbon sequestration at a rate of \$10–20 U.S. dollars per ton of carbon sequestered. Some of these have established new forests. Others have simply prevented the deforestation that would have occurred without the project, counting this as carbon gained.

B. Ocean Sequestration

Sequestration in the ocean has been discussed, but currently our scientific knowledge of marine biogeochemistry is not adequate to predict the consequences of any large-scale actions that may be taken. Discussions have centered on the use of iron to fertilize the ocean, following the demonstration in the IronEx project that the phytoplankton in the mid-Pacific are limited by iron and that they "bloom" spectacularly when a solution of iron is experimentally added from a ship. Most of the carbon fixed in photosynthesis would be rapidly released by respiration, but dead biota and feces would sediment and contribute to the biological pump, thus reducing the outgassing of CO_2. There is a secondary effect of stimulating the growth of phytoplankton. Some of these organisms produce dimethyl sulfide, which is chemically transformed in the atmosphere to aerosol sulfuric acid particles that form condensation nuclei for clouds. Thus, stimulating the phytoplankton might increase cloudiness over the ocean, which in turn would increase the reflectivity of the planet and cause a cooling effect. Although some people have been enthusiastic about the prospects of managing the carbon cycle in this way, it would be folly to attempt large-scale ocean fertilization in the absence of better knowledge of the functioning of marine ecosystems and the working of the biological pump.

C. Geological Sequestration

Sequestration of carbon in geological reservoirs is being explored. Technologies rely on the availability of a stream of CO_2 which would otherwise be dispersed in the atmosphere because the cost of concentrating CO_2 *ab initio* would be high. Target reservoirs include depleted oil and gas fields, aquifers, and the floor of the ocean. For example, in a process to recover methane from oil wells and deep coal seams (so-called coalbed methane), a stream of CO_2 is used to flush out the methane. Instead of releasing it to the atmosphere, it can be sequestered in methane-bearing coal seams in which it is apparently stable. For example, it is estimated that the San Juan basin in the southwestern United States has a total capacity of about 2 GtC, and costs are said to be comparable to the cost of sequestration by forests. There are clear attractions to geological sequestration for oil and gas companies where appropriate geological strata exist near existing pipelines and a stream of CO_2 is available.

VI. WILL THE SINKS INCREASE, REMAIN THE SAME, OR DECREASE?

In this final section, I raise questions that have not been addressed but which now seem to be urgent research questions.

1. What is the future of the terrestrial sink? Our understanding of the limits of the carbon sink is poor. In particular, it is difficult to predict the response of the sink over long periods of time, especially in future atmospheres of elevated CO_2. It is clear that terrestrial photosynthesis responds to elevated CO_2 in the range 360–700 ppm, and that in most cases growth responds to twice-normal CO_2. However, at what CO_2 concentration can this response be expected to saturate? It seems likely that nutrient supplies ultimately will become limiting in terrestrial ecosystems in the way that they are now limiting in the ocean. Models of plant growth suggest that climate warming will, in many cases, increase ecosystem respiration to a point where it exceeds photosynthesis, thus turning the sink into a source, and field observations suggest that this may have happened already in tundra ecosystems.
2. Species changes and acclimation to CO_2: To what extent will ecosystems acclimate to elevated CO_2 and temperature by species changes and reorganization, which may enable them to retain their sink strength? Answers to this question are likely to be suggested by studies of effects of past climates on vegetation, but current climatic changes are occurring at a much faster rate than before, and therefore the past may not (in this case) be the key to the future.
3. How secure is the soil carbon? The carbon content of the soil may be susceptible to climate warming and also possibly to oxidation on being disturbed by cultivation. The behavior of those pools which turn over slowly is not well understood, nor is the transport of organic material from the soil to the rivers and ocean.
4. How stable is the ocean? Questions of stability are important in the ocean carbon cycle because disturbance of the circulation patterns could result in large releases of stored carbon to the atmosphere or large

changes in uptake at the surface. It is thought, for example, that an increase in the rate of downwelling at the start of the last glacial epoch caused additional flux from the atmosphere. It is also thought by some scientists that the gulf stream, which brings warm water from the tropics to northern Europe, may have shifted substantially in the past. We do not know the stability of these circulations.

Very few authors have attempted to make predictions of global CO_2 concentrations and temperatures over periods of more than 50 years because the uncertainties are too great and technologies advance rapidly. An historical perspective illustrates this point. One hundred years ago in London, the environmental issue of the day was the inexorable accumulation of horse feces in the streets, hardly a matter of concern 50 years later in the post-Ford era. Attention then switched to atmospheric pollution, known as smog, caused by burning high-sulfur coal. Smog caused death and disease; but as a result of legislation the hazard was removed.

The track record of the scientific community in forecasting the future is not good. Scientists who have predicted "doom and gloom" in recent decades have generally been proved wrong. For example, the Club of Rome used a state-of-the-art computer model in 1972 to forecast the resource limitations of human population growth, and they forecast an abrupt downturn in the quality of life by 2000. However, economic and population growth are still occurring, despite their predictions to the contrary. They failed completely to predict the major environmental concerns about global warming that we are now discussing at the start of the new millennium.

See Also the Following Articles

ATMOSPHERIC GASES • BIOGEOCHEMICAL CYCLES • DEFORESTATION • ENERGY FLOW AND ECOSYSTEMS • MARINE ECOSYSTEMS • NITROGEN AND NITROGEN CYCLE • PHOTOSYNTHESIS, MECHANISMS OF

Bibliography

Berner, R. A. (1993). Palaeozoic atmospheric CO_2: Importance of solar radiation and plant evolution. *Science* **261**, 68–70.

Dixon, R. K., Brown, S., Houghton, R. A., Solomon, A. M., Trexler, M. C., and Wisniewski, J. (1994). Carbon pools and flux of global ecosystems. *Science* **263**, 185–190.

Falkowski, P. G., and Raven, J. A. (1997). *Aquatic Photosynthesis.* Blackwell, Oxford.

Grace, J., *et al.* (1995). Carbon dioxide uptake by an undisturbed tropical rain forest in southwest Amazonia, 1992 to 1993. *Science* **270**, 778–780.

Indermühle, A., *et al.* (1999). Holocene carbon-cycle dynamics based on CO_2 trapped in ice at Taylor Dome, Antarctica. *Nature* **398**, 121–126.

Lloyd, J. (1999). The CO_2 dependence of photosynthesis, plant growth responses to elevated CO_2 concentrations, and their interaction with soil nutrient status II. *Functional Ecol.* **13**, 439–459.

Malhi, Y., Baldocchi, D. D., and Jarvis, P. G. (1999). The carbon balance of tropical, temperate and boreal forests. *Plant Cell Environ.* **22**, 715–740.

Meybeck, M. (1993). Natural sources of C, N, P and S. In *Interactions of C, N, P and S Biogeochemical Cycles and Global Change* (R. Wollast, F. T. Mackenzie, and L. Chou, Eds.), pp. 163–190. Springer, Heidelberg.

Post, W. M., Emanuel, W. R., Zinke, P. J., and Stangenberger, A. G. (1982). Soil carbon pools and world life zones. *Nature* **298**, 156–159.

Schimel, D. S. (1995). Terrestrial ecosystems and the global carbon cycle. *Global Change Biol.* **1**, 77–91.

Schlesinger, W. H. (1997). *Biogeochemistry*, 2nd ed. Academic Press, San Diego.

Tans, P. P., Fung, I. Y., and Takahashi, T. (1990). Observational constraints on the global atmospheric carbon dioxide budget. *Science* **247**, 1431–1438.

United Nations (1992). Earth Summit Convention on Climate Change, United Nations Conference on Environment and Development, Rio de Janeiro, Brazil.

WBGU (Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen) (1998). *Die Anrechnung biologischer Quellen und Senken im Kyoto-Protokoll: Fortschritt oder Rückschlag für den globalen Umweltschutz. Sondergutachten 1998.* WBGU, Bremerhaven, Germany.

Whittaker, R. H., and Likens, G. E. (1975). The biosphere and man. In *Primary Productivity of the Biosphere* (R. H. Whittaker and G. E. Likens, Eds.), pp. 305–328. Springer, Berlin.

CARNIVORES

Hans Kruuk
Institute of Terrestrial Ecology

GLOSSARY

basal rate of metabolism The minimum amount of energy spent by an adult animal that has not eaten recently, at normal body temperature during rest (usually sleep).
eutherian Mammal in which the embryo is attached to the mother by a placenta
fissipedia suborder of the Carnivora with divided toes.
guild Species in a community that use similar resources.
marsupial Mammal without placenta and with a pouch to carry the young.
monophyletic Derived from a single ancestor.
pinnipedia Suborder of the Carnivora with fin-like limbs (seals, sea lions, and walrus).
plantigrade Walking on the soles of the feet.

CARNIVORES are a highly varied group of mostly closely related species. This article discusses the diversity of terrestrial carnivores, excluding the seals but including some marsupial species. Most of the terrestrial carnivores belong to one single order: Carnivora. Some of the taxonomic and evolutionary relationships are discussed as well as the social organizations, effects in ecosystems, and conservation status.

I. SPECIES DIVERSITY

Among the mammals that are broadly referred to as Carnivores there is a weasel of approximately 45 g and a polar bear of up to 700 kg, approximately 15,000 times heavier. There are species living almost permanently in water (sea otter), in trees (palm civet), or in deserts (fennec fox); some eat buffaloes, some eat beetles, and some eat bamboo. It is not surprising, therefore, that with such large variation and divergent adaptive features there is disagreement among researchers about the evolution and classification of species. Classification is based on morphological evidence (dental characteristics, anatomy of the skull base, and other morphological features) and on molecular genetic information.

This article will deal with the approximately 230 Fissiped carnivores, i.e., the terrestrial species of the order Carnivora and excluding the seals, sea lions, and walrus (Pinnipeds). It will also discuss a group of ecologically rather similar species of marsupials in Australia belonging to the families Dasyuridae and Thylacinidae.

The Carnivora are a monophyletic order, descended from the family Miacidae approximately 60 million years ago in the Palaeocene. The order has two main branches, the dog-like Canoidea and the cat-like Feloidea. The 33 seals and sea lions and walrus belonging to the Pinnipeds are sometimes included in the order Carnivora and sometimes given separate status. They evolved from the Canoidea, but there is disagreement about which terrestrial carnivores are their closest relatives, the main candidates being the bears and the mustelids.

The Carnivora are generally divided into seven families, although some taxonomists recognize more. The Canoidea include the Canidae (dogs) with 35 species, the Mustelidae (martens) as the largest carnivore family with 67 species, the Ursidae (bears and pandas) with 9 species, and the Procyonidae (raccoons) with 15 species. The Feloidea comprise the Felidae (cats) with 35 species, the Hyaenidae with 4 species, and the Viverridae (mongooses and genets) with 66 species. The main taxonomic disagreements are over the Ursidae and the Viverridae. The two species of panda are often thought not to belong to the bears but to deserve a separate family or they may be included with the Procyonidae, and the Viverridae are often divided into two families—the Herpestidae (mongooses, 31 species) and the Viverridae (genets and civets, 35 species).

The dogs and foxes (Canidae) constitute a highly monomorphic family. All are very dog-like with nonretractile claws, and all are coursing predators, such as wolves, coyote, jackals, foxes, wild dog, bush dog, and maned wolf. All species have much in common in their ecology and social behavior, although they may vary from solitary to gregarious. Sizes vary between that of the large gray wolf (up to 80 kg) to that of the tiny fennec fox of little more than 1 kg. Canids occur on all continents, and with the dingo they even fielded an early introduction in Australia. The gray wolf is the ancestor of all domestic dogs.

Eleven canids (31%) are endangered or vulnerable, and all the large species are threatened in at least some parts of the world (the two wolves, African wild dog, and the Asian dhole).

The members of the marten family or Mustelidae occur in both the New and the Old World, with weasels, martens, mink, polecats, skunks, otters, badgers, and others. There are relatively few species in Africa, perhaps because of competition with the similar viverrids. In Britain, there are three times as many mustelid species as members of all other carnivore families combined, and in the United States there are also more mustelids than others. Almost all are quite small animals (45 g to 45 kg), with the largest species being some of the 13 species of otters. They are usually slimly built, some with retractile claws, and with a very distinctive bouncing gait, but the family is quite variable and some species are stocky, such as the badgers. Also, in feeding and social behavior they vary greatly, from solitary predators on mammals to group-living animals feeding on earthworms or fish.

Stoats and weasels have been introduced on several islands outside their normal range, and the North American mink is now an abundant exotic in many places throughout Eurasia and South America. Some of the mustelids are almost extinct (e.g., the black-footed ferret and the European mink), some were exterminated recently (sea mink), and several are in trouble (otters) or have just returned from the brink of extinction (such as the largest of them all, the sea otter). However, only seven species (10%) are listed as endangered or vulnerable.

The bears and pandas (Ursidae) are comparatively large and some are huge (weighing up to 700 kg), stockily built, plantigrade with nonretractile claws, and very short tails. The exception is the very aberrant red panda (a small arboreal species with a long tail), a species often classified as a procyonid. All but the polar bear are mostly vegetarian; they are solitary animals with a fairly simple social organization. The bears evolved more recently than the other carnivores and successfully colonized areas as far apart as the drift ice in the Arctic and the dense forests of the Old World and neotropics. Bears of colder regions hibernate. More than half of the species (five or 56%) are endangered or vulnerable, and all are in trouble in at least part of their geographic range.

Members of the raccoon family Procyonidae are all rather long-bodied animals, relatively small (up to 8 kg), with long, ringed tails, and they walk plantigrade with nonretractile claws. They occur naturally only in the New World. In addition to the various species of raccoon throughout the Americas, there are also other Neotropical species, such as the coatis, the kinkajou, and the cacomistle. They are carnivorous and insectivorous as well as herbivorous. Their social behavior is quite variable; most are solitary but some, for example, female coatis, live in large packs. Also, common raccoons may occupy winter dens in groups. None are classified as endangered, although some raccoon species are known from only one or two islands, on which they may be rare. In several areas of Eurasia common raccoons are now frequent as an introduced species.

The Viverridae are a large Old World family of small animals, occurring almost entirely in Africa and Asia. They are the most "primitive" carnivores, i.e., the closest relatives of the carnivores' original forebears, the Miacidae. The family includes many species of genet and civet, linsang, binturong, a large range of mongooses, meerkat, fossa, fanaloke, and others. Some taxonomists recognize the mongooses as a separate family—the Herpestidae (otherwise classified as a subfamily Herpestinae). The mongooses are long-bodied and look somewhat mustelid-like, with nonretractile claws; in contrast, the rest of the viverrids look and walk much more cat-like, with retractile claws. The fossa from Madagascar is unusually large (up to 20 kg), but all other Viverrids are much smaller, especially the mongooses. Most are solitary, but some mongooses and meerkats live in large packs. Most of their food is invertebrate, but they also take small mammals, etc. and occasional vegetable food.

Few viverrid species appear to be endangered (four or 6%); of these four, three are on Madagascar, an island beset by deforestation problems.

The Hyaenidae have only four species left, despite a fossil record showing large numbers in the geological past, and they occur only in the Old World. Three are relatively large species of 50 kg or more and are typical coursing animals with a dog-like build and nonretractile claws. Hyaenids are carnivorous; they often scavenge, although some also take vegetable food, and the smallest, the aardwolf, is a termite specialist. The largest species, the spotted hyena, is a gregarious hunter of large mammals, but the others are solitary. One of the four, the brown hyena, is classified as vulnerable as a species, but in many countries the others are also endangered.

Finally, there are the Felidae, the proper cats, a highly monomorphic family with many species. All are built as stalking predators, with retractile claws, and they range in size from fairly small to large, from serval and flat-headed cat and many other small species to lynx and tiger (up to more than 300 kg), various leopards, cheetah, and lion (up to 250 kg). All are carnivorous, but small species also eat many invertebrates. The social organization of all species is very similar—solitary and territorial, with the lion as the only group-living exception. Felids occur naturally on all continents except Antarctica and Australia. The wild cat *Felis lybica* has been domesticated and it has been introduced virtually everywhere in the world; it is now the most widespread carnivore.

There is concern about the conservation status of almost all cat species, especially the ones with desirable fur. Twelve (34%) are endangered or vulnerable, including all the large ones such as tiger, cheetah, jaguar, and various leopards, but not the lion.

The mammalian carnivore families outside the Carnivora proper are the Australian marsupials Dasyuridae and Thylacinidae. There are four dasyurid species that are very similar to Carnivora: three quolls or, as the Australians call them, "native cats" and the Tasmanian devil. They are relatively small, with the quolls similar in size and appearance to martens or mongooses and the Tasmanian devil more like a small badger in size and shape, and they are typically carnivorous, solitary, nonterritorial animals. The thylacine was the only species in its family and it was also the largest of these marsupial carnivores (comparable in size to a large jackal or coyote). It was exterminated very recently. The other, dasyurid species now have very much reduced geographical ranges and they are vulnerable.

II. SIZE AND ECOLOGY

There is a large diversity of sizes among carnivores, and the ecological implications of this variation are considerable. The basal rate of metabolism of carnivores tends to be greater than that of same-sized herbivores, and a large predator expends more energy than a small one; therefore, it has to capture more prey. However, the increase in metabolism with size is not linear, and when expressed as energy requirements per kilogram of body weight a large predator is more efficient than a small one (Fig. 1).

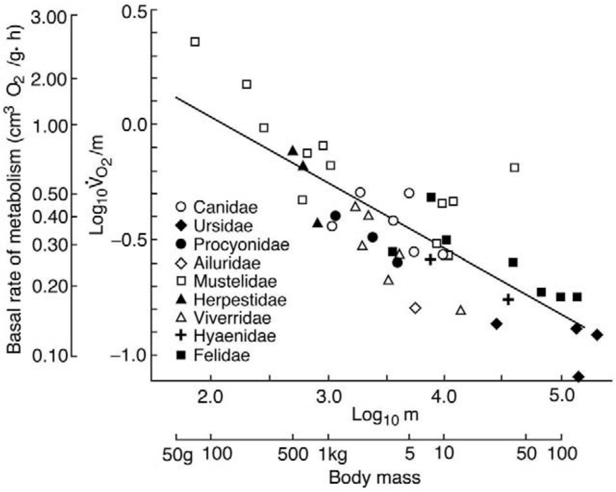


FIGURE 1 Basal rate of metabolism (oxygen consumption per gram per hour at rest) in carnivores as a function of body mass. Reprinted from B. K. McNab, "Basal rate of metabolism, body size, and food habit in the order Carnivora," in *Carnivore Behavior, Ecology, and Evolution* (J. L. Gittleman, ed.). Copyright © 1989 Cornell University. Used by permission of the publisher, Cornell University Press.

The body mass of a carnivore, when combined with the group size in which it lives, is correlated with the size of its home range, although there is much variation (Fig. 2). Much of this residual variation is due to effects of habitat and to differences in diet (predators of vertebrates having larger ranges than insectivorous ones). There is no significant correlation between body mass and gregariousness, i.e., large carnivores are no more likely to live in groups than are small ones.

Brain size of carnivores increases with body size. Irrespective of this, a meat-eating, hunting carnivore or an omnivorous species generally has a relatively larger brain than an insectivorous one. Bears have relatively large brains, whereas Viverrids and Hyaenids have relatively small ones and the others have intermediate-sized brains.

In general, large carnivores take larger prey than do small ones, but there are important variations. First, some of the largest (bears) are mostly vegetarian. Second, the trend varies between carnivore families. Within the mustelids there is a nonsignificant negative correlation between predator body mass and prey size, so the predator–prey size principle does not apply. Both felids and canids show a strong positive correlation, but the slope of the relationship is different: The increase in prey size with predator body mass is much steeper in felids (Fig. 3).

There are many specific deviations from the previously discussed general patterns, but the trends are significant. Greater energetic demands on larger carnivores are met by a dependence on larger prey, which is likely to make the large predators more vulnerable. Many of the larger carnivores are threatened globally or in parts of their range, e.g., many of the large spotted cats and the tiger, wolves, and hyenas.

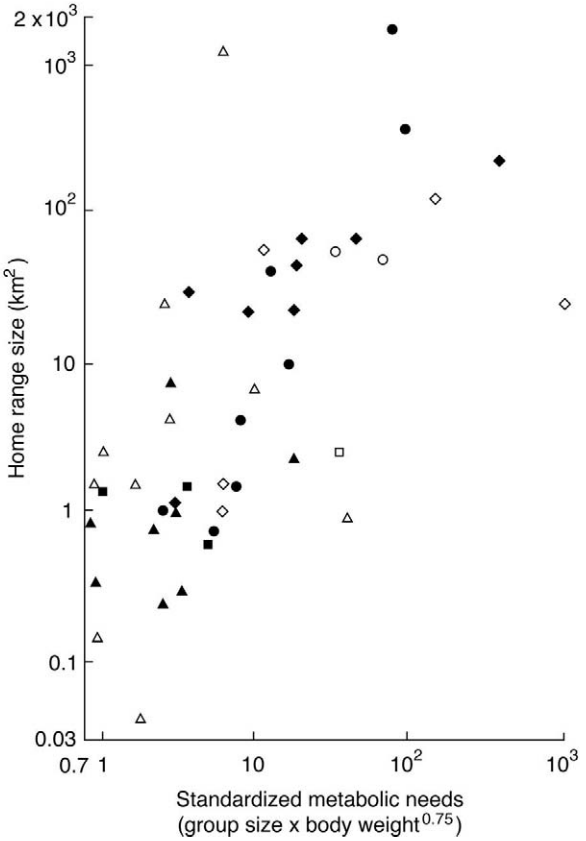


FIGURE 2 Home range size of carnivores as a function of metabolic needs. ●, Canidae; ○, Ursidae; ■, Procyonidae; □, Ailuridae; ▲, Viverridae; △, Mustelidae; ◆, Felidae; ◇, Hyaenidae (reproduced with permission from Gittleman and Harvey, 1982. Carvinore home-range size, metabolic needs and ecology. *Behav. Ecol. Sociobiol.* **19**, 57–63, Fig. 1, © Springer-Verlag).

III. FORAGING AND PHYLOGENY

The diet of carnivores in general, and that of their marsupial equivalents in Australia, consists of vertebrate prey but also invertebrates and vegetable matter. The order Carnivora shows dental specializations enabling easy digestion of vertebrates, such as large canines and the carnassial shear, but in several species the molars have been further adapted to a more grinding function when eating plants.

Foraging or hunting behavior of carnivores consists of variations on a general theme. A search leads to detection and selection of a potential victim or other source of food, and it is followed by an approach which may contain elements of stalking and/or chasing. The actual capture of prey may include seizure, immobilizing, and killing, and this is followed by eating, taking food to cubs, or sometimes quietly caching it for later consumption. Parts of this sequence may be absent, and in fact most carnivores will just search and then eat small food items without much further ado. However, even if there is a full-blooded hunt, some species never stalk (e.g., dogs or hyenas), others never chase (e.g., cats), and some may not show any specific killing behavior but just eat (e.g., hyenas).

The chain of events aimed at the capture of prey may be broken off at any stage depending on circumstances; it can also be started at any stage. It is highly adaptive, depending on prey and environment, with the predator's own motivation (its degree of hunger) apparently affecting especially the early searching stages of the hunting sequence.

Conspicuous in the predatory behavior of many carnivores is the phenomenon of surplus killing, i.e., kill-

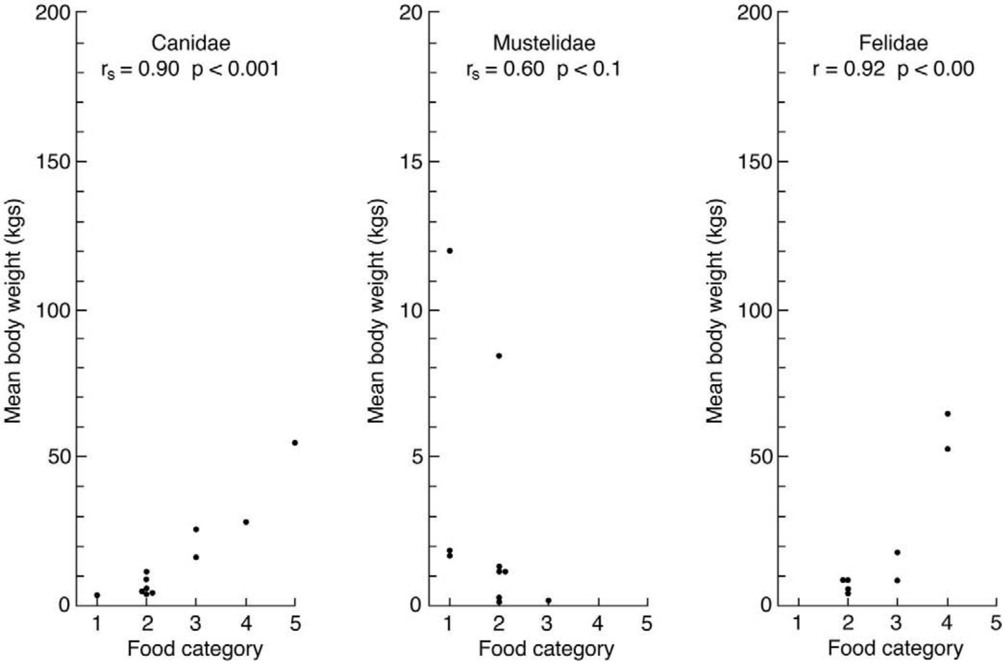


FIGURE 3 The relationship between size of predator and prey. For each species of carnivore, mean body weight and main food category are given. 1, invertebrates; 2, small rodents; 3, rabbits and hares; 4, larger mammals less than 50 kg; 5, mammals more than 50 kg (reproduced with permission from Kruuk, 1986).

ing more than is required for immediate consumption. Classic examples are the fox in the hen house or gull colony or a lion among a cattle herd; it has been described for hyenas, polar bears, wolves, leopard, and others. Large numbers of animals were killed without being eaten. All these situations have in common a lack of defense by the prey: The prey may be immobilized by particular weather conditions, it may be penned in, or it may have lost its antipredator defense through domestication. In these cases, the predator is sated and no longer hungry, but hunger normally affects only the early stages of the carnivore hunting sequence, especially the search. If, for some reason, no search or stalk or chase is needed because the hunter suddenly finds itself close to the quarry, then the rest of the hunting sequence is put in train irrespective of hunger, and there is no inhibition to capturing and killing. Functionally, such events are wasteful from the carnivore's point of view because they reduce prey availability without the predator getting the benefits.

Food caching (i.e., the storage of food in a hidden place) is a behavior pattern that limits this waste and utilizes the consequences of surplus killing. Many species do it, and do it in many different ways. In all canids it is highly stereotyped: A small hole is dug, and one single prey is dropped into it and covered with earth or vegetation by sweeps of the snout. In other families there are many variations on the caching theme: Leopards take a carcass high into a tree; spotted hyenas cache chunks of food in shallow water; brown and striped hyenas push it into a dense bush; stoats, mink, and other mustelids may make large stores by dragging numbers of prey into a single hole; wildcats may put remains of their quarry under a log; and some of the larger cats may civer a carcass with vegetation. The methods are consistent within the species, for canids within the family, and for martens within the subfamily, suggesting that caching has evolved in carnivores on several different occasions. There is no evidence of caching for the procyonids or viverrids, nor is it done by the marsupial carnivores.

Although food caching may use some of the surplus kills, it still does not utilize all the apparent waste. Not only are some of the caches never revisited by the perpetrator but also in the larger surplus kills only a small proportion of the victims are stored. There may be scores of dead gulls left by a fox and of gazelle left by spotted hyenas. In some carnivores caching is particularly highly developed. For instance, foxes are able to remember where they stored what, and they return preferentially to the more desirable cached items. Foxes also use some kind of bookkeeping system for their caches, leaving a drop of smelly, long-lasting urine near those caches which they have emptied.

Foraging and hunting behaviors are features that differ with the phylogeny of the predator. Canids invariably are coursing and running predators, with an occasional semistalking approach of prey, and they often forage by "sniff and search." At the other extreme, felids search almost entirely dependent on vision and they approach their quarry in a highly concealed stalk or ambush. Most members of the other families, including the marsupial carnivores, show the canid sniff and search behavior pattern. Only the genets and civets have a stalking behavior that is similar to that of the felids.

When a relatively small vertebrate prey is caught, the felids, many of the mustelids, and civets will kill it by severing the spinal cord with their canines. Canids also have a specific killing method—violently shaking the prey. Felids kill large prey with a throat bite, or they suffocate it with a bite over the nose and mouth. In almost all other predator–prey interactions there is no specific killing or immobilizing behavior—the predator just eats the prey. Most species of carnivore have highly specific ways of dealing with a prey and its carcass, and it is often possible to distinguish afterwards which predator was responsible for a kill.

Some subfamilies have evolved extreme foraging specializations. The Lutrinae (otters) dive for fish and crabs using their tactile senses, but essentially their hunting is also based on the canid pattern. Several Melinae and Mellivorinae (badger) species pursue their prey by digging after it. Species such as the otters and African wild dogs use energetically demanding behavior to catch their quarry (a high investment and high reward strategy), which makes them especially vulnerable to fluctuations in prey density and to food loss to scavengers.

As a result of the variation in foraging methods between families and subfamilies, there are also phylogenetic patterns in the diet. Compared with many other species of mammals and birds, diet analysis for carnivores is relatively easy, although it has some serious problems. Food usually consists of clearly discrete items, which can be recognized and quantified. Furthermore, scats are often easy to find, no matter how elusive the animals may be, and scat analysis has become a major tool despite the difficulty of relating scat content to diet in a quantitative manner. Also, for many species it has proved possible to obtain direct, quantitative observations of predation and foraging behavior. The result is an extensive body of knowledge of carnivore diet, the important link in the relation between the predators and their environment.

One summary showed that of 111 species of carnivore (from all families), only 36% could be classified as predominantly meat eaters, i.e., taking more than 60% of their diet in the form of other mammals or birds. Indeed, in that analysis among representatives of families such as the bears, raccoons, and viverrids, there were no proper meat eaters at all, and they were found to feed on insects, vegetation, or a mixture of various food categories. Many species were called omnivorous if their diet did not include 60% of any one category of food. However, such a limit is quite arbitrary, and many important species were left out of the analysis because of lack of information. Therefore, it is an oversimplification, and one could also summarize the diet differently (Fig. 4): The majority of carnivore species will eat meat and will prey on other mammals at some time or other. However, the point had to be made that many other kinds of food are involved.

There is a clear importance of phylogeny in the diet: The food of a felid tends to be more like that of another felid than that of a canid and vice versa, and this applies to several families (Fig. 4). All families except the felids are intensive exploiters of vegetable and invertebrate food sources: The Felidae are the most exclusively carnivorous—the ultimate predators. Only the felids, hyaenids, and canids feed significantly on large mammals, with an occasional exception in the other families. The viverrids, which are probable ancestors of felids and hyaenids, do not kill larger mammalian prey but they are either insectivorous or have a very mixed diet. The diets of the four hyenas are more different from each other than between the members of the other families, and their specializations range from wildebeest to termites or melons or carrion. The bears and pandas are vegetarians. Species in the raccoon family all have a mixed diet, including much vegetable matter. Canids have mostly varied diets, composed of insects, fruits, and mammals. Many of them are proper meat eaters, but even these often include some vegetable matter, quite unlike the cats.

Some of these family-specific trends are further refined in the subfamilies. The mostly meat-hunting mustelids also include two subfamilies with 9 badgers, which almost all feed on invertebrates and vegetation, and a subfamily of 13 otters that subsist on fish or crabs.

Not only are there differences between families and subfamilies in the kinds of prey or vegetation which they select, but also the degree of specialization varies (Fig. 5). Specialization can affect an animal's vulnerability to environmental change. Among the canids, each species uses an average of 6.5 (out of 10) major prey categories, each constituting at least 1% of its food. On the other end of the scale, each bear only uses 3.7 prey

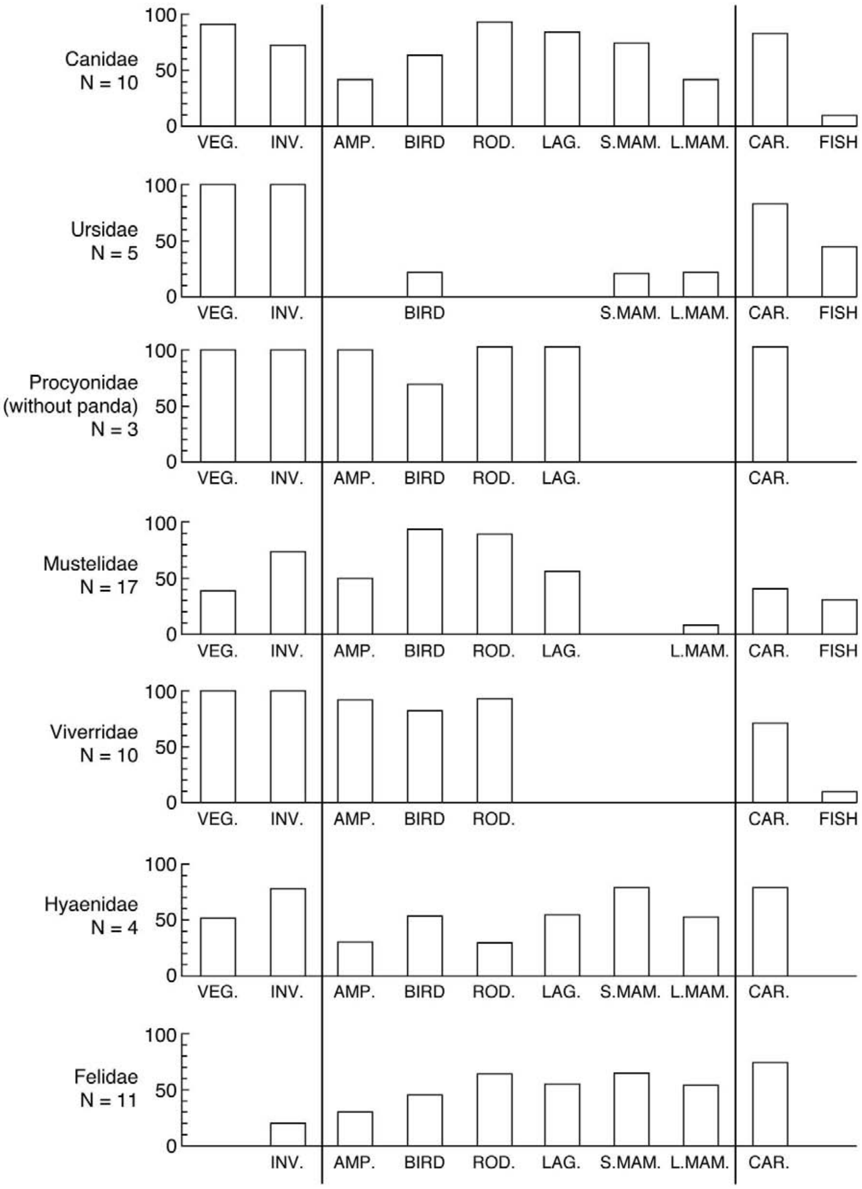


FIGURE 4 The use of different foods by species from different families. Percentage of species in each family in which a particular food category constitutes more than 1% of diet. Veg, vegetable foods; inv, invertebrates; amp, amphibia and reptiles; rod, small rodents; lag, rabbits and similar size prey; s.mam, large mammals smaller than 50 kg; l.mam, large mammals larger than 50 kg; car, carrion (reproduced with permission from Kruuk, 1986).

categories and each felid 4.0, so they are much more specialized. The other carnivore families are intermediate. This specialization (dependence on a few resources) may make bears and felids more vulnerable to environmental change than canids.

The degree of specialization can be described only in the broadest of terms because it is difficult to measure and quantify. Terms such as omnivore, opportunist, generalist, and specialist have no absolute values, and they refer to animals which may be selecting from a variable set of availabilities. There is also a problem with definitions. For instance, the Eurasian badger is highly focused in its food selection in any one area, concentrating entirely on earthworms in northwestern

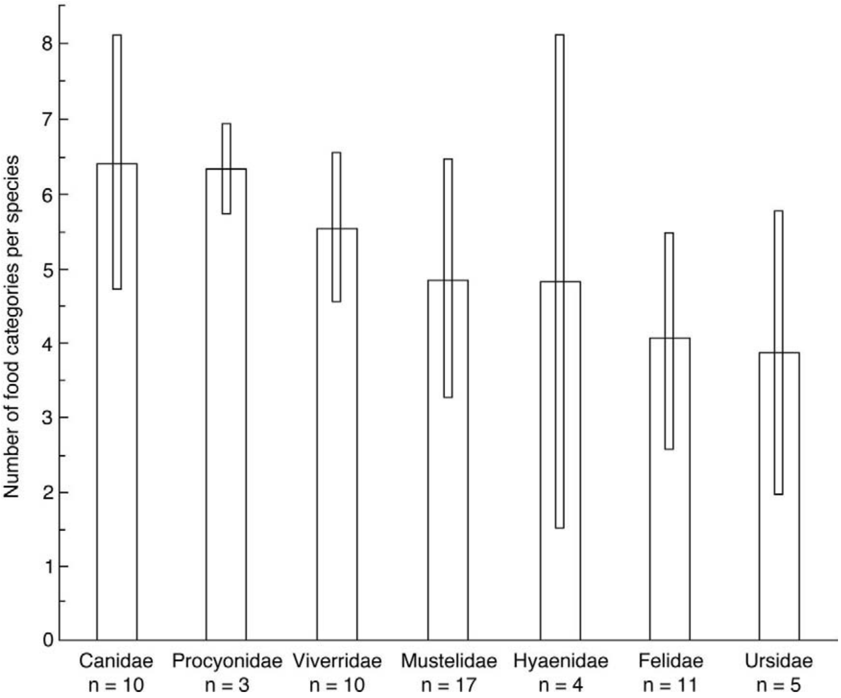


FIGURE 5 The number of food categories used by species of different carnivore families. *N*, number of species studied, with mean and standard deviation of the number of major food categories (see also Fig. 4) (reproduced with permission from Kruuk, 1986).

Europe, on rabbits in southern Spain, and on olives in northern Italy. There is no doubt that in each of these areas badgers are highly specialized, but their specializations are different in different places and overall they could be described as opportunists. Currently, there is no good quantitative descriptive term for such a pattern.

Nevertheless, despite inadequate terminology it is recognized that in any one place some species use many more different prey categories than others. For instance, a cheetah on an African savanna takes almost only small or medium-sized ungulates on open grasssland, whereas a leopard in the same area is much more catholic in its tastes. It eats the same mammals but also much smaller mammals as well as birds and snakes, stalking them in the open and in dense bush or between the rocks. Along European streams a mink will eat small mammals, frogs, fish, birds, and insects, whereas along the same banks the (larger) otters feed almost exclusively on fish and frogs, clearly much more focused. Such a broad comparative indication of specialization suggests the dependence of a predator on few or many prey categories, even though there may be difficulties with labeling as specialist or opportunist.

In general, because closely related animals often have similar food habits, food selection may be termed a conservative characteristic in the evolution of carnivores. There are many exceptions; for instance, among the meat-hunting felids there is a fishing cat, and the hyena family includes the aardwolf, which feeds on nothing but termites, the spotted hyena, which is an exclusive large ungulate hunter or scavenger, and the striped and brown hyenas, which are as catholic in their tastes as possible. However, these exceptions do not invalidate the overall importance of ancestry in the animals' environmental relationships.

IV. CARNIVORE GUILDS IN ECOSYSTEMS

Closely related members of the same carnivore family tend to exclude each other, although there are important and puzzling exceptions. Of course, competition between natural populations of different carnivores is rarely observed, probably because those cases in which it occurred have long ago come to their natural conclusion, i.e., the demise of one of the contestants. However, when perturbations occur competition may be obvious.

For instance, the introduction of the American mink in Europe totally removed the European mink from most of its range. The famous wolf population of Isle Royal arrived there in the early 1950s, and it completely replaced a population of coyotes. There are several observations of wolves killing coyotes, apparently without the coyotes being eaten. In many other areas in North America, just one of the two species is found, although usually these places would appear to be suitable for both. Coyotes tend to replace red foxes.

Red foxes are aggressive to arctic foxes and exclude them when their areas overlap, but arctic foxes can feed and survive at much lower temperatures. Recent red fox increases (e.g., in northern Scandinavia) caused the demise of arctics over large areas. In the 1970s, when the black-backed jackals of the Serengeti were decimated by disease, the very similar and previously rare side-striped jackal population increased dramatically. Tigers are reported to often exclude leopards, and different species of otters exclude each other.

Also, species that are not closely related but have overlapping ecological niches may affect each other. For instance, the African wild dog disappeared from the Serengeti between the 1960s and 1990s probably partly because of an increase in populations of hyenas and lions.

Exceptions to this pattern of exclusions are the three species of African jackal. They are very similar and closely related; nevertheless, their geographical ranges overlap considerably and they can be seen eating from the same carcass.

There are other examples of perturbations or species introductions causing the disappearance of competing carnivores, but such events are relatively rare. Thus, it is often assumed that populations of predators each have their own ecological niche and rarely affect each other. However, the ecosystems that we see are end results, and often what we see may be the status quo long after earlier populations have been wiped out or have been prevented from moving in.

One way in which predator species adapt to an ecological niche in a particular habitat is through morphological or behavioral variation. When potentially competing species share a range, differences between them (e.g., in size) may be greater than those from animals that live in ranges not shared with the other species. This is known as character displacement. One of its consequences is that in a guild of predators in any one area a character, such as body size or size of the canines, shows a fairly evenly spaced stepwise distribution among species (Fig. 6). The obvious effect of this is to minimize competition. Such divergence of morphological characters does not always occur, however, and its striking demonstration in British mustelids is contrasted with an almost absence in Serengeti jackals. The marsupial carnivores of Tasmania also show a clear character divergence.

The existence of such structured mechanisms, to avoid competition within predator guilds in ecosystems, suggests that resources for the carnivores may be at a premium. This in turn presents the possibility that the predators depress prey populations.

It is often argued that predators have little or no effect on prey numbers, as demonstrated by ecosystems such as the Serengeti with more than 25 carnivore species that coexist with the many prey in apparently stable populations. However, such an apparent stability is the result of interactions over a long period of time, and it is possible that many prey species were extinguished in the past and many predator–prey relationships have

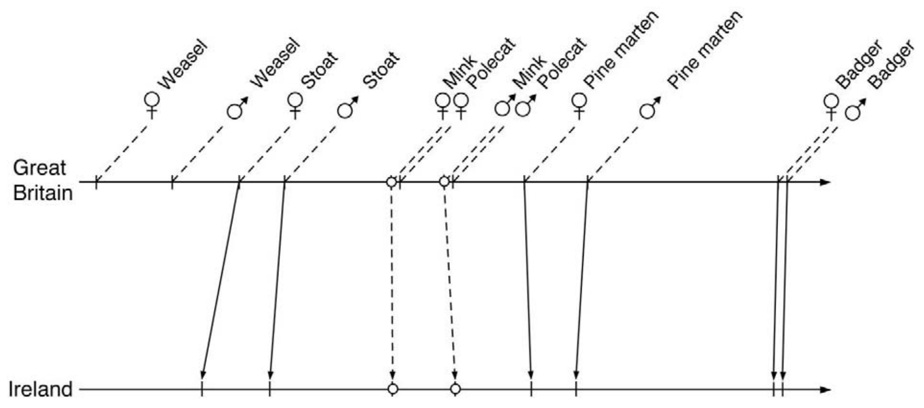


FIGURE 6 Even spacing of mean lengths of skulls of Mustelid species in Britain and Ireland, log scale. The mink has recently been introduced (reproduced with permission from Dayan and Simberloff, 1994).

not had a chance to become established (as first arrivals in the ecosystems were eliminated).

The deliberate or accidental introduction of carnivores into ecosystems in which they were previously absent provides an opportunity to demonstrate effects of predation. Introduced predators (cat, dog, red fox, mongoose, American mink, stoat, and ferret) have caused extinctions and declines of endemic prey species throughout the world, including many birds, mammals, and reptiles. There is little doubt, therefore, that carnivores cause extinctions or prevent populations from becoming established, and they must have done so in the evolutionary past. On all continents the fauna would look very different from what it is now if there had been no carnivore predators.

Also in the predator–prey systems which currently exist, the effects of carnivore predation can be far-reaching. Predator removal often dramatically increases numbers in prey populations compared with those in suitable control sites, as demonstrated by foxes preying on partridges. Pest species such as rabbits can be limited in numbers by carnivores operating in conjunction with other mechanisms such as disease. Wolves may exercise major effects on populations of caribou, moose, and white-tailed deer, and spotted hyenas can have a major effect on wildebeest. Often, other limiting factors are involved simultaneously, e.g., predators may be the means whereby a herbivore population is maintained at the carrying capacity of the vegetation.

V. CARNIVORE SOCIAL SYSTEMS

The majority of carnivores are solitary and territorial. Their spatial organization is maintained by females (with or without offspring) defending a range against other females of the same species, and by males defending a (larger) range against other males. Male ranges may overlap with several females ranges. Territorial behavior involves scent marking and visual displays as well as direct aggression.

There are some important variations to this generalization. The marsupial carnivores are not territorial; therefore, instead of a regular spacing between individuals, they are more or less randomly distributed. This begs the question, not satisfactorily answered to date, why eutherian carnivores spend so much time and effort on risky territorial behavior if other species with similar ecology can do without.

The organization of canids is pair based rather than solitary; a male and a female share the same range. This is also the only family in which males usually assist in providing the offspring (by regurgitating food). The fact that such pair behavior is tied with the species' phylogeny suggests that in individual species it is not necessarily adaptive.

From the simple, territorial arrangement a group organization has evolved independently in several species in all carnivore families except the bears, and such group living has complicated land tenure considerably. Among the canids the most striking examples are the wolf, the wild dog, and the dhole in Asia, but in several other canid species offspring may also remain for 1 or 2 years, overlapping in time with subsequent litters of the basic pair. This has been described for several jackals, foxes and others.

African wild dogs live in extremely tight packs, often composed of 20 or more individuals, almost always close together, with on average twice as many adult males as females. The wolf has a very different organization. It also lives in sometimes large packs, but it does so in a much more "fission–fusion" type of society, with individuals coming and going, sometimes hunting in groups and sometimes alone, within the pack territory. In both species usually only one female per pack breeds.

Group organizations in other carnivore families show as many different patterns as there are species, with, for instance, lions in permanent prides of up to 20 related females joined by small groups of males which are replaced every few years. Spotted hyenas live in female-dominated clans of up to 80 in group territories, and individual members may hunt or join with others or they may be solitary within the group territory. Cubs are cared for almost entirely by their own mothers. Both banded and dwarf mongooses occur in dense packs, with the sexes mixed, and all members of the pack care for all offspring, Eurasian badgers forage on their own and look after their own cubs only, but they live in group territories defended by both sexes. Female Eurasian otters, on the other hand, may occupy individual core areas within a group territory of five or six females, whereas males remain in their own ranges.

For the phenomenon of group living in carnivores, there are no striking, simple phylogenetic trends and no substantial effects of body size, climate, prey size, predation, or various other factors to explain the occurrence of packs, bands, or prides. It does not appear that social species have done any better or worse than solitary ones in terms of numbers or densities, nor is their future survival more or less endangered. There is, however, one set of environmental factors which appear to affect gregariousness, i.e., the distribution of food or of resources in general.

Currently, there is only one general hypothesis to explain grouping of carnivores—the resource dispersion hypothesis, first explicitly presented by Macdonald (1983). In principle, it relates the size of territories and the numbers of animals inhabiting each territory to the spatial and temporal pattern of resources. For instance, it suggests that badgers in northwestern Europe live in groups because their main food, earthworms, occurs in well-spaced patches, each of which is available only at certain times. To be able to feed at all times a badger needs a large area with several patches, with the size of the area determined by the scatter of the patches. Once a badger has such a territory it can accommodate several more individuals with only limited or no competition. The hypothesis is attractive, but it does not explain all the intricacies of spatial organization. To date, there is no comprehensive alternative.

Carnivores living in groups may or may not hunt cooperatively, depending on species or conditions. Clear advantages of cooperation, in terms of hunting success or energy balance, have been demonstrated for wild dogs and spotted hyenas. Lions collaborate on a well-organized basis, with each pride member occupying a preferred place in the hunting formation and deviations of this pattern resulting in lowered success.

VI. CHANGES IN DIVERSITY

When the Carnivora first appeared, approximately 60–40 million years ago, there were other, similar animals already well established. For instance, the extinct order Creodonta included families such as the Hyaenodontidae, and there were various large marsupial predators such as the Thylacoleo or pouch lion. Proper carnivorous feeding and predation have evolved several times independently—at least twice among the marsupials (in the Borhyaenidae in South America and some Dasyuridae in Australia) and twice among placental mammals (in Creodonta and Carnivora).

The large carnivorous expansion from the Palaeocene onwards coincided with the evolution of angiosperms, flowering plants and grasses (evolving away from the ferns and gymnosperms such as conifers). This resulted in a large floral diversification, including savanna-type vegetations, and enabled the extensive diversity of ungulates and rodents to evolve. This in turn enabled the evolution of specialist predators.

Most of the other, noncarnivore large predators have now gone extinct, with the last creodonts occurring approximately 8 million years ago and the last really large marsupial carnivores occurring 2 million years ago, when species of Homo were already well established. Of the smaller marsupial carnivores, just a few small dasyurids still occur in Australia. However, while the other predators slowly disappeared, the order Carnivora diversified into a multitude of different families, genera, and species.

Why the creodonts disappeared, while at least initially carnivores thrived and probably replaced the creodonts, is a mystery. Creodonts and carnivores were closely related and the skeletal remains, such as vertebrae and the locomotory system, were similar. The difference between creodonts and carnivores was not much greater than the variation within these groups. However, obviously there is more to an animal than its skeleton, and the reasons for extinction may well have lain in other aspects of morphology, physiology, or behavior.

In the early stages of evolution the advantages appear to have lain with the creodonts, and only after 20 million years did the balance tip in favor of the Carnivora. The Carnivora have been four times as successful as the creodonts: The latter are known from 45 genera spanning approximately 45 million years, whereas Carnivora, excluding living and aquatic genera, are know from 218 genera over a span of more than 55 million years.

Carnivore species are also fewer now than they were in the geological past. In fact, most of the Carnivora have become extinct, and although there is still a rich complement of species, there were many more in the past. For instance, we know of 333 genera in the seven extant families of carnivores, of which 237 (71%) are extinct. Many complete carnivore families have also disappeared, just like the creodonts and large marsupial predators.

An interesting phenomenon was the occurrence of saber-tooth species. Saber teeth evolved several times independently in different species, families, and even orders and included the Megantereon, Homotherium, and Machairodus, which were at least as big as a lion, and formidable felids such as the North American sabertooth cat Smilodon. Some of these were present at the same time as early species of Homo, but all of them are extinct. They pose a difficult problem to paleontologists. Large canines are used by today's carnivores for killing prey and for social purposes such as fighting opponents over territorial claims. However, were the extra-large saber teeth—the huge, flat daggers which were seemingly far too large for any jaw—used to kill extra-large prey, for opening carcasses, or what? In fossil assemblages it was always the very largest ones, the top predators, which sported saber teeth.

There is no likely explanation for saber teeth in the acquisition of food. The fragile, sharp weapons, often with serrated inner edges, must have been quite useless against thick skin or on large bodies, with the gape of the owner being insufficient to use these canines effectively. However, saber teeth were obviously effective weapons; otherwise, they would not have evolved several times independently. However, they also disappeared again in all these cases. It is likely that saber teeth made use of some Achilles heel in their prey (of which we have no evidence today), but that in response the prey species evolved means of protection. It was an arms race which was eventually lost by the sabers, but we do not know who conquered and why.

Homo is a highly successful hunter, and mankind was and is in competition with Carnivora at several different levels. Homo also preys on and is preyed on by Carnivora. Therefore, are people the cause of carnivore extinctions? The question is often asked, and answers are far from straightforward. Many carnivore species extinctions occurred approximately 4 million years ago, at the time that hominids arrived on the scene, and many followed during the next 2 million years.

However, many extinctions occurred before man arrived, and at the time Homo's arrival was not the only event which changed the environment. There were drastic changes in climate, for instance, approximately 3.2, 2.4, and 0.8 million years ago. The Pleistocene Period started at 0.8 million years ago with a massive climatic shift that coincided with the appearance in Eurasia of African species such as the lion, leopard, spotted hyena, and perhaps the major (although not the first) movement out of Africa of Homo. If such major dispersal events occurred in conjunction with climatic changes, it is equally likely that extinctions would have occurred.

For many of the carnivore extinctions, we cannot blame our own species with any conviction. Homo may have been closely involved or not at all, and perhaps environmental change rendered species more vulnerable to competition and predation by mankind. Of course, recent extinctions have been fairly well documented, and here mankind's guilt is in no doubt. The Tasmanian wolf or Thylacine was exterminated by sheep farmers in the 1920s and 1930s. The North American sea mink was obliterated in the late nineteenth century for its fur, and the sea otter almost followed it into oblivion; it was barely saved and it has recovered fairly well. The "wolf" of the Falkland Islands was still there when Charles Darwin visited, but sheep farmers exterminated it and since the 1880s it has existed only in museums. Several carnivores were totally eradicated from Britain, including the brown bear and the wolf, whereas wildcat, polecat, and pine marten have only just managed to survive. There are long lists of carnivore extinctions from throughout the world, and there is no doubt that most of these were man induced.

Whatever caused the demise of carnivore species in the past, it is important to prevent it from occurring in the future. Several species are on the brink of extinction in the wild, including tiger, panda, European mink, and black-footed ferret, and many more face local extermination. In geological terms, the diversity of carnivores is decreasing extremely rapidly.

See Also the Following Articles

FOOD WEBS • MAMMALS, BIODIVERSITY OF • PREDATORS, ECOLOGICAL ROLE OF

Bibliography

Creel, S., and Creel, N. M. (1995). Communal hunting and pack size in African wild dogs, *Lycaon pictus*. *Anim. Behav.* **50**, 1325–1339.

Dayan, T., and Simberloff, D. (1994). Character displacement, sexual dimorphism, and morphological variation among British and Irish mustelids. *Ecology* **75**, 1063–1073.

Gittleman, J. L. (Ed.) (1989). *Carnivore Behavior, Ecology and Evolution.* Cornell Univ. Press, Ithaca, NY.

Gittleman, J. L. (Ed.) (1996). *Carnivore Behavior, Ecology and Evolution*, Vol. 2. Cornell Univ. Press, Ithaca, NY.

Gittleman, J. L., and Harvey, P. H. (1982). Carnivore home-range size, metabolic needs and ecology. *Behav. Ecol. Sociobiol.* **19**, 57–63.

Kruuk, H. (1986). Interactions between Felidae and their prey species: A review. In *Cats of the World: Biology, Conservation and Management* (S. D. Miller and D. D. Everell, Eds.), pp. 353–373. National Wildlife Federation, Washington, D.C.

Kruuk, H. (1989). *The Social Badger.* Oxford Univ. Press, Oxford.

Kruuk, H. (1995). *Wild Otters: Predation and Populations.* Oxford Univ. Press, Oxford.

Macdonald, D. W. (1983). The ecology of carnivore social behaviour. *Nature London* **301**, 379–384.

McNab, B. K. (1989). Basal rate of metabolism, body size, and food habits in the order Carnivora. In *Carnivore Behavior, Ecology and Evolution* (J. L. Gittleman, Ed.), pp. 335–354. Cornell Univ. Press, Ithaca, NY.

Mills, M. G. L. (1990). *Kalahari Hyaenas.* Unwin-Hyman, London.

Stander, P. E. (1992). Cooperative hunting in lions: The role of the individual. *Behav. Ecol. Sociobiol.* **29**, 445–454.

Woodroffe, R., Ginsberg, J., and Macdonald, D. (1997). *The African Wild Dog.* IUCN, Gland, Switzerland.

CARRYING CAPACITY, CONCEPT OF

Gregg Hartvigsen
State University of New York, College at Geneseo

GLOSSARY

density dependence The condition that environmental factors influence population growth rate in relation to population size. Density dependence usually is seen as an linear, inverse relationship between population growth rate and population density (i.e., population growth decreases as density increases) and may occur if individuals compete or predators are more effective as a prey population increases.

density independence The absence of environmental factors that influence population growth as a function of density. This may occur if mortality removes a fixed percentage of a population, independent of population size.

logistic growth Regulated population growth that follows the logistic equation $dN/dt = rN(1 - N/K)$. Populations growing according to this equation increase rapidly at low densities and the growth rate decreases as they approach carrying capacity (K).

population A group of individuals of a particular species that live in a region. A population is usually a subset of the entire species.

population regulation The constraint of positive population growth. The study of population regulation deals with the factors that cause this constraint, such as competition for food or predation.

population stability The tendency for populations to return to a previous size after a disturbance, such as reductions due to hunting or disease or increases due to immigration. Stable populations may be locally stable (return after small disturbances) or globally stable (return after severe or catastrophic disturbances).

CARRYING CAPACITY is the maximum number, density, or biomass of a population that a specific area can support. This number is likely to change over time and depends on changes in environmental factors (e.g., rainfall and temperature), resources (e.g., food, hiding places, and nesting sites), and the presence of predators, disease agents, and competitors over time. The concept of carrying capacity has been explicitly recognized for approximately 150 years and its use has waxed and waned during this time. Currently, the use of carrying capacity to describe any particular population is made only with great caution, although the concept remains intuitive and fosters questions that address our fundamental understanding of what factors regulated populations over time and space.

I. INTRODUCTION

Populations, or groups of individuals within a species, change over time. There is general agreement among ecologists that population growth is bounded by biotic and abiotic environmental factors that result in approximate, maximum numbers of organisms that can be supported in different habitats. A population's carrying capacity is difficult to measure and likely varies over time and through space. The concept of carrying capacity has played an important role in the fields of basic ecological research, wildlife management, and conservation biology.

The concept of carrying capacity also involves determining how many people Earth can support. Dialogues about human carrying capacity are often quite contentious and illustrate the difficulty surrounding the concept of carrying capacity.

The concept of carrying capacity is alive and well, although some have argued that it should be abandoned altogether. On the surface, the concept is easy to understand and intuitive and, therefore, is likely to stay with us for some time. However, recent developments in our understanding of the dynamics of population change over time have greatly modified what is considered an area's carrying capacity. Therefore, we need to recognize both the strengths and the weaknesses of the concept of carrying capacity.

II. ORIGIN OF THE CONCEPT OF CARRYING CAPACITY

Humans have long been aware of the limitations of their own population growth. As early as the Old Testament one can argue that the concept of population limitation was recognized. In the book of Genesis (28:3) are the following words: "And God Almighty bless thee, and make thee fruitful, and multiply thee, that thou mayest be a multitude of people." Although rather vague, as quoted from the King James version, a "multitude" in English is defined as "a great number of people" (*Merriam-Webster Dictionary*) that eventually would be spread over the earth and would not be an ever-increasing population. The number at which a population reaches and remains sustainable is referred to as the "carrying capacity."

Recognition of carrying capacity probably occurred long before written history began. It is likely that the earliest agriculturists, perhaps 10,000 years ago, were keenly aware of the number of mouths that an area could sustainably feed and that increasing numbers of people required increases in food production (i.e., area of land in cultivation). Long before agriculture, hunter-gatherer groups likely were aware of the sustainable number of members that regions could support, although it may be argued that high mortality rates inhibited these early populations from pushing the limits of sustainability. During difficult times populations reached or exceeded what we might think of as a carrying capacity, possibly imposing nomadic lifestyles in which groups had to intermittently move after local resources were depleted. Speculating on past population dynamics hints at the trouble with the concept of carrying capacity: It must represent a dynamic value that changes over time and is highly dependent on many interacting factors, such as environmental variability and, for early humans, what hungry animals awaited their forays. This makes the concept difficult to use.

Charles Darwin, on page 116 in his 1859 book *Origin of Species,* quoted in his third chapter on the struggle of existence from Malthus' 1798 "An Essay on the Principle of Population as It Affects the Future Improvement of Society":

> "A struggle for existence inevitably follows from the high rate at which all organic beings tend to increase. Every being, which during its natural lifetime produces several eggs or seeds, must suffer destruction during some period of its life, and during some season or occasional year, otherwise, on the principle of geometrical increase, its numbers would quickly become so inordinately great that no country could support the product. Hence, as more individuals are produced than can possibly survive, there must in every case be a struggle for existence, either one individual with another of the same species, or with the individuals of distinct species, or with the physical conditions of life. It is the doctrine of Malthus applied with manifold force to the whole animal and vegetable kingdoms; for in this case there can be no artificial increase of food, and no prudential restraint from marriage. Although some species may be now increasing, more or less rapidly, in numbers, all cannot do so, for the world would not hold them."

Malthus' clear recognition of the importance of limitations of growth in populations helped Darwin to lay the foundation for his theory of natural selection, which is built on the premise that populations are regulated primarily by competition which leads to differential reproduction.

The earliest concise description of carrying capacity derives from Pierre Francois Verhulst, a Belgian who lived in the mid-nineteenth century. Verhulst, perplexed by accounts that the human population appeared to be increasing exponentially, derived a mathematical formula which he called the "logistic" equation that would account for a slowing in the population growth rate as a function of population size. The same relationship was rediscovered in 1920 by Raymond Pearl and Lowell Reed, who used the logistic equation to predict the population of the United States based on census data collected from 1790 to 1910 (apparently neglecting the fact that the area of the United States increased more than three-fold during this time). The resulting application of the logistic equation to U.S. census data led Pearl and Reed to greatly underestimate the U.S. population, predicting it would level off at about 197 million in the Year 2050 (Fig. 1).

The failure of Pearl and Reed to accurately predict the population of the United States, currently at about 275 million, reveals at least one important aspect of mathematical models. It indicates that, for the U.S. human population, some aspect of the logistic equation must be wrong. Two candidate problems include violations of the assumptions that the area of the United States remained constant and that the U.S. population lacked immigration. These differences between data and model predictions can help us to better understand the problem at hand. In this example, the violations of model assumptions leads us to re-evaluate the factors that influence population growth and ultimately regulate it. The prediction from the model also clearly did not hold for a variety of socioeconomic and human health reasons and may have been wrong for statistical reasons. Using the logistic equation (based on the filled data points in Fig. 1) to predict population size well into the future, without providing confidence limits, is clearly tenuous at best. These problems have led us to be more careful in our predictions of how populations change over time and whether a carrying capacity can be predicted from such data or even exists.

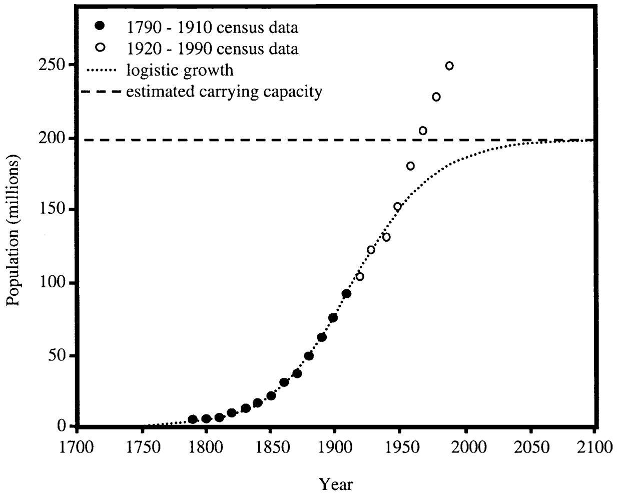


FIGURE 1 Data on the growth of the U.S. population from 1790 to 1910 (●) and from 1920 to 1990 (○). The dotted line represents Pearl and Reeds' fit of the logistic equation, yielding a carrying capacity (*K*) of about 197 million people, estimated to be reached in about the Year 2050. The population in the Year 2000 is about 275 million (not shown).

III. DEFINITIONS OF CARRYING CAPACITY

Carrying capacity is the maximum population that a given area can sustain. The measures commonly used include the number of individuals or the total biomass of a population, which are each highly dependent on differences in physiology and age structure among species and across large taxonomic groups. The use of the term carrying capacity has changed over time, but most models suggest that population growth is rapid when density is low and decreases as populations increase toward some maximum. In addition, any definition of this concept improves as we narrow the time and area for the population that we are studying. Population descriptions, therefore, are often depicted as densities, accounting for the number of individuals per unit area. Population density usually varies over time and from place to place. In practice, we generally use population size or density to describe carrying capacity, which is determined either by resource availability or by the influence of enemies (predators and/or pathogens).

Various definitions of carrying capacity arose in the twentieth century, ranging from the suggestion that carrying capacity is that level below which predators have no effect on a population to the population size which can be maximally supported in a given region (previously referred to as the "saturation level"). There also has been a distinction made between "ecological carrying capacity," which refers to the limitation of a population due to resources, and a management-oriented, maximum sustainable yield for a population, referred to as an "economic carrying capacity," which is usually lower than ecological carrying capacity. These definitions clearly lead to difficulty for wildlife managers who have been preoccupied with attempting to determine whether populations are either too high or too low. These debates continue, as exemplified by range management decisions in Yellowstone National Park and issues regarding the increasing frequency of reintroduction programs of top predators.

Carrying capacity may best be expressed mathematically. One of the simplest forms of population change over time can be represented as the differential equation $dN/dt = rN$, where dN/dt represents the instantaneous change in a population over a short time period, r is the intrinsic growth rate of the population, and N is the size of the population. This yields what is often referred to as a "J" curve, or exponential growth (Fig. 2). In discrete time this relationship is referred to as geometric growth.

In 1838, Verhulst modified the exponential growth equation and derived the logistic equation that depicted population growth rate as being inversely related to population size. To slow population growth he added an additional term yielding $dN/dt = rN(1 - N/K)$, where K is the population carrying capacity. The term "$1 - N/K$" slows growth rate linearly toward zero as the population (N) approaches the carrying capacity (K). This results in a sigmoidal S-shaped curve for an increasing population over time (Fig. 2). If the population exceeds K ($N > K$), then $1 - N/K$ is negative, causing growth rate dN/dt to be negative and the population to decline monotonically toward K.

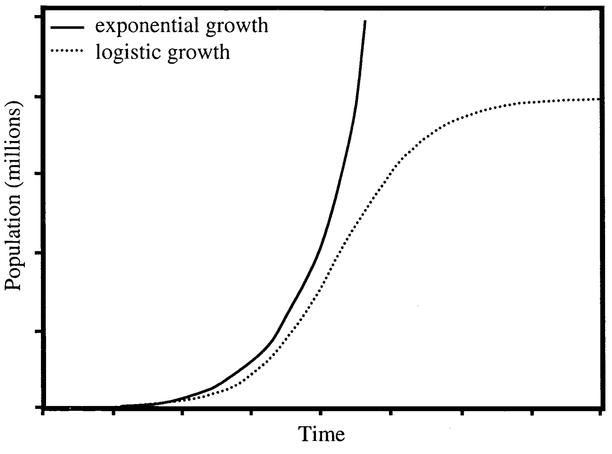


FIGURE 2 Comparison of unregulated exponential growth (solid line) with regulated logistic growth (dotted line).

An important attribute to bear in mind is that the logistic equation is deterministic, meaning that if we use the equation to predict population size at the end of a fixed amount of time we will derive the same population each time we start the population over. This assumption is usually violated in field conditions in which random effects, such as accidental deaths, failure to find mates, or fluctuations in environmental conditions, are common. Therefore, it has been argued that we should not expect real populations to behave according to the logistic equation.

This simple equation has been challenged repeatedly by critics without apparent damage. This resilience of a theory is rather rare in science, which is a discipline that prides itself on being able to quickly dispel hypotheses (or equations) given even a small amount of contradictory data. However, the intuitive nature of the idea that populations are regulated by factors such as food supply helps the logistic equation to remain a staple in ecological texts and classrooms. The reason this equation and carrying capacity (K) endure is that the equation's shortcomings help us better understand the dynamics of real populations, ensuring its utility for many years to come.

The discrete, or difference, form of the logistic equation yields a different prediction of population behavior compared to the previously described continuous version. In particular, the discrete form was the equation used by Sir Robert May to first describe how a simple, deterministic equation could produce chaotic population dynamics, a pattern that emerges when intrinsic growth is relatively high. This chaotic behavior appears to mimic realistic changes in populations over time. Several long-term data records conform to chaotic dynamics, including the change in the number of lynx captured over time in Canada (Fig. 3).

IV. DO POPULATIONS HAVE CARRYING CAPACITIES?

This question has been addressed using a variety of techniques, including observational types of studies that rely on long-term time series data sets such as the number of lynx captured over time (Fig. 3), highly controlled laboratory experiments (Fig. 4), and mathematical models to determine potential mechanisms through time series reconstruction. The short answer

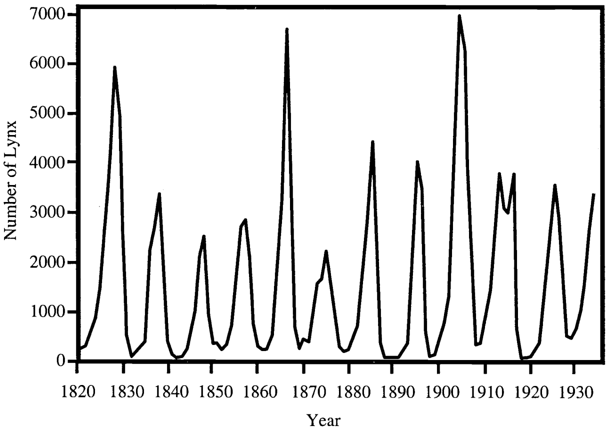


FIGURE 3 The number of lynx trapped in the Mackenzie River district from 1821 to 1934 (after Elton and Nicholson, 1942).

is that there likely are carrying capacities for most species but that determining these at any one point in time and space is quite difficult.

Many controlled laboratory experiments have been published that show populations behaving in a fashion consistent with the logistic model (i.e., populations reach a carrying capacity). One of the earliest studies was completed by Gause in 1934 (Fig. 4). Most laboratory populations tend to increase and then reach some level at which they fluctuate around what might represent a carrying capacity. It is interesting to note that populations rarely exhibit a smooth transition between a growth phase and gradual reductions in growth rate ending in stable populations, despite the controlled environmental conditions. An additional caveat to consider is that prior studies that concluded population growth patterns differed significantly from the expected logistic growth likely suffered disproportionately during the review process and failed to make it into print.

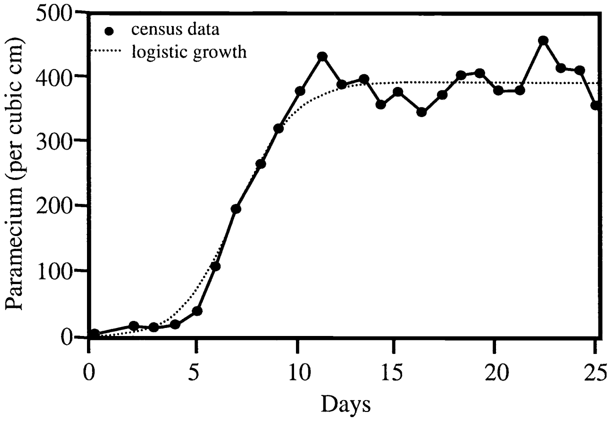


FIGURE 4 The change in the density of *Paramecium caudatum* over time in the laboratory. The dotted line represents the best fit logistic equation (after Gause, 1934).

Determining how natural populations change over time is surprisingly difficult. The first requisite information necessary to determine whether a population is at or near a constant size, considered here to be a proxy for the habitat's carrying capacity, is simply the population's size over time. This often has to be determined over long periods of time in ways that are accurate, reliable, and repeatable. In field studies, it is rare to have the luxury to repeatedly estimate population size, a technique allowing us to assess the accuracy of our estimates. Determining that a population fluctuates may represent real changes in populations or represent either natural variability (statistical "errors") or actual errors in our estimates.

Assuming we overlook these shortcomings in our data, what do populations do? In general, populations usually fluctuate over time. We may be able to correlate these changes with biotic or abiotic factors or some function of the two with time. Sometimes the fluctuations cannot be distinguished from random noise. Some populations, including the classic examples of lynx, hares, and lemmings, cycle periodically (Fig. 3). The persistence of population cycles over long periods of time has led to great speculation regarding the factors that might lead to periodicity. Recent work suggests that simple causal mechanisms of cycling are unlikely and that a combination of random environmental factors and nonlinear, density-dependent factors influence populations.

V. DETERMINING THE CARRYING CAPACITY OF A POPULATION

Determining the carrying capacity of any particular population at a particular time is not trivial. Many different techniques have been suggested and tested, including three primary techniques that can be used to attempt to detect a change in population growth rate as a function of population size: mathematical modeling of specific mechanisms, tactical experimental tests in the laboratory and field, and statistical analysis of time series data. Ultimately, a combination of these techniques will enable us to understand the importance of regulation in populations and the degree to which populations appear to be governed by a carrying capacity. I have already discussed the importance of the logistic equation and will briefly introduce the empirical approaches.

There has been much interest in and work completed to determine what factors regulate the change in growth of populations over time. The factors that slow population growth rate change over time and from location to location and differ for different species. Regulating factors also are likely to interact with each other, thus complicating the determination of a population's carrying capacity.

In a classic study, Davidson and Andrewartha in 1948 used a partial regression technique to analyze an experiment designed to test the relative influence of biotic and abiotic factors on regulating a small herbivorous insect population. They concluded that 78% of the population variance was due to abiotic or weather-related factors. In particular, the number of individual thrips in the spring was related mostly to the preceding autumn climate. This study was influential because it provided strong evidence that this population of thrips was regulated not by biotic factors such as competition or predation but rather by abiotic factors.

A second method used to detect the presence of density dependence on population regulation is the analysis of time series data. The best data are those that have been collected over consecutive years and that exceed the periodicity of both observed environmental and population fluctuations (generally >10 years). These data can be subjected to tests that investigate the relationship of change from year to year as a function of the population during the previous year or years in order to detect whether the population appears to be regulated. Such analyses, however, are unable to provide any information on the underlying mechanisms that might lead to population regulation. Therefore, time series analysis is an excellent exploratory tool that can be used to investigate the possibility that a population is regulated. This information can then suggest experiments designed to partition variance among potential candidate mechanisms.

Determination of a population's carrying capacity is best done through a combination of modeling, experimentation, and time series analysis. Research efforts, however, need to be directed toward investigating the underlying mechanisms that govern population regulation. Without an understanding of the relative importance of these regulating factors, it will be difficult to determine whether populations are regulated and whether we can detect a population's carrying capacity.

VI. CURRENT RESEARCH ON CARRYING CAPACITY

Two main areas of research continue to drive our quest to understand population regulation and the strength and importance of carrying capacity. The persistence of these questions indicates the need to clarify the mechanisms that influence population change over time.

A. Determining the Relative Strengths of Factors That Regulate Populations

Although some researchers have argued that populations are unregulated, most agree that negative feedback mechanisms operate on populations, resulting in decreased growth at high densities. This may occur through changes in the abundance of food, through increased predation or disease, or through a combination of these biotic factors and abiotic factors such as local climate. These factors may reduce birth rates or increase death rates, or both. Although there are circumstances in which these rates change at low population densities (e.g., the Allee effect, which states that very small populations are likely to decrease due to such factors as difficulty in finding mates or pollen limitation), their regulation at high densities is likely to be common. This change in birth and death rates as a function of density is referred to as "density dependence."

A population that is regulated has intrinsic, extrinsic, or a combination of these factors that slows population growth. Under such conditions a population's per capita growth rate decreases with increasing population size through reduced birth rates and/or increased death rates. This relationship, in logistic growth, is assumed to be linear. The existence of a carrying capacity, however, is not dependent on the shape of this function, so the violation of this linearity assumption does not weaken the concept of carrying capacity. A better understanding of this relationship, generally determined through carefully designed experiments, will help us understand the importance of regulation on population dynamics.

Krebs *et al.* (1995) suggested that hare and lynx cyclic population dynamics are likely influenced by different sets of factors, including food availability and predation driving the dynamics of hare populations and the lynx population is driven primarily by changes in the number of hares. In a more highly controlled experiment using three trophic levels, Hartvigsen *et al.* (1995) determined that plant performance was controlled by the interaction of top-down and bottom-up factors, including the level of plant resource availability and the presence or absence of herbivores and herbivore predators. These studies suggest that complex, interacting biotic and abiotic factors likely influence population dynamics.

B. Determining Population Carrying Capacity

The logistic growth equation attempts to model regulated population change over time and relies on several important assumptions, including the absence of time lags (population dynamics is independent of prior events), migration or immigration, genetic variability or selection, population age structure, and the fact that density dependence is linear (each individual added to the population has a similar effect on the population's per capita growth rate). Violations of these assumptions have been found in various populations and have led to more refined, realistic, and complicated forms of the logistic equation. In addition, the model assumes that carrying capacity (K) is constant over time and space. This assumption occasionally may be valid in situations in which a population is regulated by habitat availability. This might occur, for example, where the number or area of nesting sites is fixed. It is easy to conjure up situations, however, when this assumption would be violated over very small spatial or temporal scales. It is not likely that K would be constant since populations are usually limited by resources, competitors, enemies, and often combinations of these factors that vary with the environment over time. Under these conditions changes in resource availability can influence population size directly or indirectly through its often nonlinear effect on the population of competitors and/or predators.

In addition, there is great concern about the stability and persistence of threatened and endangered species (see Section VII). Work in this area has begun to recognize the importance of species interactions, immigration and emigration among subpopulations, the introduction of exotic species, and other factors that violate the assumptions of simple logistic growth. The movement of individuals among subpopulations enables the possibility of increased long-term persistence of populations by reducing large-scale fluctuations and the spreading of risk that a species will become extinct in the event that a single local population disappears (becomes extirpated). This area of research, referred to as "metapopulation biology," involves determining long-term viability of these subdivided species, and there is currently much research activity in this area.

VII. THE IMPORTANCE OF THE CONCEPT OF CARRYING CAPACITY TO BIODIVERSITY

The concept of carrying capacity suggests that species are likely to have some upper limit to their population. If the upper limit is "hard," then we expect populations to achieve this state and remain relatively constant. Populations, however, as demonstrated in Figs. 3 and 4, do not behave in such a simple fashion and have rather "soft" limits. As such, populations usually exhibit random, cyclic, or chaotic dynamics. These dynamics generally lead to increased chances that populations will reach the absorbing state of zero (become either locally extirpated or globally extinct).

We must be concerned about the dynamics of relatively small populations over time. The probability that a population will go extinct is generally related to the degree to which it fluctuates (population amplitude and frequency). Therefore, processes that cause populations to increase fluctuations are likely to lead to species loss and associated reductions in biodiversity. Thus, conservation efforts may be needed that will buffer populations and associated habitats from extreme fluctuations. Conservation efforts are often directed toward increasing a population's carrying capacity. It should be kept in mind, however, that constant environments also may lead to species losses. The intermediate disturbance hypothesis has gained much empirical support and suggests that the maximum number of species that an area can support occurs when disturbances are intermediate in either frequency or impact. We should be concerned that our management efforts do not reduce the carrying capacity of target species.

VIII. THE HUMAN CARRYING CAPACITY

The best estimates of human population indicate that it has continued to grow exponentially over recorded history, although the actual growth rate has changed over time. Attempts to fit data on the human population to the logistic equation have failed (Fig. 1), and current indications are that no human carrying capacity can be predicted from simple population statistics. However, we might ask whether our population growth rate is likely to slow down in the foreseeable future and, ultimately, reach a stable carrying capacity or whether it will overshoot its carrying capacity and eventually collapse.

Joel Cohen (1995) found that estimates of the human carrying capacity have ranged between 1 billion and 1 trillion people, with the majority of estimates falling between 4 and 16 billion (the current population is about 6 billion). These estimates suggest that we are approaching an apparent limit for our species. Regardless of which estimate seems most appropriate as an upper limit for humans on Earth, the growth rate of our population will eventually slow to ≤ 0. This can occur as a result of increasing death rates and/or decreasing birth rates. I predict that as our population grows in the coming decades there will be an increase in mortality due to diseases. The effect of disease agents on controlling population growth will likely increase due to increases in human contact rates and rapid transit, increasing evolution of drug resistance, and increasing virulence rates. These factors also may reduce birth rate, which of course presents a more pleasant alternative to slow population growth.

Can we avoid a population crash? I venture the guess that we cannot. Any long-term stabilization of the human population will require a decrease in the current global birth rate. We certainly cannot hope to achieve a relatively stable population without invoking a substantially higher death rate than the current rate, which is not a comforting thought. It is difficult to imagine, however, that the influence of disease will operate in a simple density-dependent fashion. Instead, it seems more plausible that diseases will "break out" more often with increasing population size and with larger scale consequences, bringing about a strong reduction in our population—a response seen in many other populations that have increased beyond their carrying capacities.

One last area of hope is that individuals will lower their consumption rates, thereby adjusting the human carrying capacity. It is unlikely that Earth can support tens of billions of people with lifestyles matching those of people in the developed nations such as the United States. Therefore, there remains a chance that changes in human behavior will allow our population to gently transition toward a sustainable, zero growth rate population.

IX. CONCLUSIONS

The concept of carrying capacity has a history that spans at least thousands of years. The formal definition is about 150 years old and is generally coupled to the asymptotic population in the logistic growth equation (K). Critics argue that because of ongoing confusion and the multitude of definitions attached to the concept we would be better off to simply abandon the term. We also must be concerned that the term not be used to advance any particular political agenda associated with determining how large populations of any particular species "should" be in particular areas. This entry cautiously suggests that the concept remains useful. Since most populations are likely to be at least occasionally limited by factors that depend on the population's density, we need to continue advancing our knowledge of how populations behave and use this information to guide the design of laboratory and field experiments aimed at determining the mechanisms that regulate populations. Only by using the combination of field and laboratory techniques, grounded in a theoretical framework that has roots going back to the simple logistic equation, will we hope to understand and conserve populations, including our own, over long periods of time.

See Also the Following Articles

POPULATION DENSITY • POPULATION DYNAMICS • POPULATION STABILIZATION (HUMAN) • POPULATION VIABILITY ANALYSIS (PVA) • SUSTAINABILITY, CONCEPT AND PRACTICE OF

Bibliography

Caughley, G. (1979). What is this thing called carrying capacity? In *North American Elk: Ecology, Behavior and Management* (M. S. Boyce and L. D. Hayden-Wing, Eds.). The University of Wyoming, Laramie.

Cohen, J. E. (1995). *How Many People Can the Earth Support*? Norton, New York.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection*. Murray, London.

Davidson, J., and Andrewartha, H. G.(1948). The influence of rainfall, evaporation, and atmospheric temperature on fluctuations in the size of a natural population of Thrips imaginis (Thysanoptera). *J. Anim. Ecol.* **17**, 200–222.

Dhondt, A. A. (1988). Carrying capacity: A confusing concept. *Acta Oecol.* **9**, 337–346.

Elton, C., and Nicholson, M. (1942). The ten-year cycle in numbers of lynx in Canada. *J. Anim. Ecol.* **11**, 215–244.

Gause, G. F. (1934). *The Struggle for Existence*. Macmillan, New York.

Hartvigsen, G., Wait, D. A., and Coleman, J. S. (1995). Tri-trophic interactions influenced by resource availability: Predator effects on plant performance depend on plant resources. *Oikos* **74**, 463–468.

Krebs, C. J., Boutin, S., Boonstra, R., Sinclair, A. R. E., Smith, J. N. M., Dale, M. R. T., Martin, K., and Turkington, R. (1995). Impact of food and predation on the snowshoe hare cycle. *Science* **269**, 1112–1115.

Pearl, R., and Reed, L. J. (1920). On the rate of growth of the population of the United States since 1790 and its mathematical representation. *Proc. Natl. Acad. Sci. USA* **6**, 275–288.

Turchin, P. (1999). Population regulation: A synthetic view. *Oikos* **84**, 153–159.

CATTLE, SHEEP, AND GOATS, ECOLOGICAL ROLE OF

Andreas Troumbis
University of the Aegean

GLOSSARY

grazing intensity Frequency and closeness of grazing.

grazing pressure Stocking rate, the units of grazing animals per land area.

overcompensatory growth In grazed plants, the reaction to tissue removal by enhanced primary production compared to in undefoliated controls.

THE INTERACTION OF HUMANS WITH DOMESTICATED RUMINANTS has strongly influenced the cultural evolution of societies. Livestock breeding has both negative and positive effects on landscape and ecosystem processes and biodiversity. Ecological disturbances generated by domestic herbivores are defined at various spatial and temporal scales, from the micro-scale of the defoliation of an individual plant, to nitrogen redistribution at the pasture level, to global-scale phenomena such as desertification and toxification of the planet. This article focuses mainly on the interactions of cattle, sheep, and goats with other biotic components of grazed ecosystems.

I. DOMESTICATION OF RUMINANTS: THE PROCESS THAT CHANGED THE FACE OF THE EARTH

The natural capacity of ruminants to digest hard plant tissue and to transform it into animal biomass is a service that has been provided by biodiversity to human societies for millennia. The cultural evolution of humans is strongly related to the exploitation of this service, which has influenced—and in turn has been influenced by—the pace, the pattern, and the distribution of societies. The first traces of domestication have been found in the Iranian mountains, where there are paleolithic settlements of sheep and goat herders dating from ca. 10,000 B.P., suggesting that domestic ruminant keeping precedes proper agriculture. Evidence of organized livestock breeding culminates between 8000 and 6000 B.P., at several sites in the Middle East and Anatolia. Ali Kosh, located at the head of the Persian Gulf in the ancient kingdom of Susiana (7000 to 6500 B.P.), is such a site where villages of farmers-breeders are found. Animals—and plants—were originally domesticated in a specific region, and then diffused during a 10,000 year process to many other parts of the world.

In the past, nomadic tribes and their herds of domesticated animals roamed freely to avoid drought, in search of good grazing land. In the rangelands of North America, Native Americans followed the wandering herds of buffalo, and it was not until domesticated cattle and sheep were introduced by European colonists that the vegetation and soils of the plains and mountains were put under stress. The rise of national boundaries has created a barrier to many such movements. Seasonal transhumance is still a practice in Mediterranean areas, where especially flocks of sheep are moved between winter quarters in the mild rainy lowlands and summer quarters in rich alpine pastures.

On a global scale, diffusion was drastically accelerated by European expansion from the fifteenth century onward. At present, rangelands cover a large part of the world's land surface (40%). Eighty percent of these rangelands are in arid and semiarid areas, where most of the 50 million humans who are directly dependent on livestock breeding for their subsistence live. In Africa, livestock breeding—mostly cattle and goats—is the main source of income south of the Sahara, especially in countries like Sudan, Chad, and Somalia. This is also the case in much of Asia, in particular in Mongolia, the central Asian republics of the former USSR, and Tibet. Cattle and sheep are also economically important domestic animals in New Zealand, Australia, Europe, and parts of the United States.

The domestication of animal species, especially of ruminants, and the associated land management and breeding practices are historically important ecosystem disturbance agents through the effects of grazing, trampling, and digging. They have transformed, shaped, and selected landscape and vegetation types, and have had significant, often dramatic, effect on many ecosystem properties, including plant and animal diversity. Uncontrolled grazing on shrublands and the creation of pastures have gone hand in hand with the creation of agricultural land. Native plant and animal communities have been greatly modified by the introduction of domestic animals, either directly or through habitat alteration and landscape change. There are many well-documented examples from all around the world, but those that occurred on islands are the most informative for our purposes.

The direct effects of the introduction of domestic grazers on native faunas since prehistoric times are well described for the Mediterranean islands, where original faunas have been affected by species extinction and introductions promoted by humans. The modern mammalian faunas on the big islands of the Mediterranean (e.g., Corsica, Sicily, Crete, and Cyprus) replaced the entirely endemic Holocene faunas that existed before human arrival. After the spontaneous extinctions of the late Pleistocene, the surviving endemic herbivores (e.g., *Episoriculus* spp; *Prolagus sardus, Rhagamys orthodon, Meridiopitymys henseli*) were rapidly eliminated, even before A.D. 7000, either intentionally by hunting or indirectly through the effects of commensal species. The faunas of today are the result of a severe human-induced selection in favor of species belonging to the geographic and cultural universe of the human groups that immigrated to the islands, within which domestic livestock predominate.

In southern Greenland, the arrival of the Norsemen in ca. A.D. 1000, and the introduction of cattle have disturbed the fragile dwarf-shrub heaths, which had established slowly after the deglaciation period. The activities of Norsemen, but especially of their cattle, destroyed the equilibrium between climate and vegetation, causing severe erosion of the unstable soil. When the Norsemen disappeared in the fifteenth century, a new phase of soil stabilization began, but this is now being severely disrupted with the immigration of more humans, now as sheep-breeders, in the beginning of the twentieth century.

Madagascar's highland region was once covered with evergreen forests dominated by about 20 endemic tree species. Beginning around A.D. 600, Indonesian settlers started to remove forests to create swidden fields. At about A.D. 1000, zebu cattle were introduced from Africa, further increasing the need to expand grassland at the expense of forests. These forests were being permanently replaced by a floristically impoverished steppe vegetation on ferrolitic soils. By A.D. 1600, the highland forest had mostly disappeared: tree and humus removal had led to massive erosion, floods, water shortages, and faunal extinctions or endangerment.

Although much poorer in species richness, the faunas of the French subantarctic islands (Amsterdam, Saint Paul, Kerguelen, and Crozet Islands) show an analogous history following their discovery in the sixteenth and seventeenth centuries. Sheep, mouflon, and cattle are among the nine introduced species that thrive owing to lack of competitors and predators, and despite the small number of founding individuals. Herbivores have induced particularly significant changes in the nature and the structure of plant communities, leading to the extinction of endemic plants (e.g., *Phylica nitida*) or the degradation of the fragile peat-bogs that constitute the nesting sites of the rare Amsterdam albatross (*Diomedea amsterdamensis*).

In North America's Great Basin Desert, changes in plant communities that occurred after the introduction of domestic livestock in the late 1800s resulted in unusual, unforeseen cascade effects on the interactions between native mammalian species. The establishment and dissemination of cheatgrass (*Bromus tectorum*) in the Great Basin has played a central role in this process. Evidence suggests that cheatgrass was introduced accidentally as a grain contaminant at the end of the nineteenth century, at the same time that large-scale domestic animal grazing began. Imported from Mediterranean Europe and central and southwestern Asia, seeds of cheatgrass exploited an ecological niche, as no native annual was dominant in the Great Basin. Cattle, sheep, and feral horses facilitated its establishment, for they spread the seeds in the same areas that they disturbed. Once established, cheatgrass promoted the likelihood of fire to the detriment of the native species. In addition, other factors, such as the effects of the lack of vesicular-arbuscular mycorrhizae and selective lagomorph grazing, have worked in concert to further establish cheatgrass dominance. The ecological consequences of this establishment have been an increase in fire frequency and intensity, a decrease in species diversity, and a landscape that is susceptible to severe erosion.

Ecosystem change resulted in the subsequent irruption of mule deer (*Ocodoileus hemionus*) and the expansion of mountain lions (*Felis concolor*). Domestic sheep depredation is currently increasing in western North America and is related to the expansion of suitable mountain lion habitat, and consequently of the lion's distribution and abundance. Furthermore, the expansion of this predator's population has caused the severe reduction of the porcupine (*Erethizon dorsatum*), another North American native species.

Although environmental determinism is not the dominant explanation of agricultural geography, the limits of the distributions of domestic animals, especially of their races, are often attributed from a physiological point of view to the physical environment, temperature, and water availability on a regional scale. However, it is also quite clear that the driving forces behind some of these ecological changes stem largely from cultural and social factors, including the size and growth of human populations, which determine the scale and the practice of stock-keeping activities.

II. ECOSYSTEM-LEVEL AND GLOBAL EFFECTS OF LARGE DOMESTIC ANIMALS

The ecological role of cattle, sheep, and goats, that is, the effects they have on ecosystem processes and biodiversity, is far more complex than the fact that they remove biomass by eating grass and often browse, or feed on, twigs, shoots, and leaves of other plant species. The floristic composition and productivity of the grazed ecosystem and the persistence of vegetation and of plant and animal species, as well as physical and biogeochemical processes, may suffer substantial changes because of grazing animal-related factors, such as the severity and frequency of grazing, species of animal, method of prehension, treading, excreta deposition, and even the saliva deposited on plants. In some environments, the effects of grazing are quite predictable, whereas in others such effects are much more difficult to predict. Under certain conditions, grazing can act as a completely density- and species-independent disturbance, and under other conditions the effects on plant can be very selective and effects on soil properties and chemistry very heterogeneous in space.

A. Impacts on Vegetation

The impact of domestic ruminants on vegetation is typically studied through comparisons between grazed and ungrazed ecosystems. The dependent variables in these studies are descriptors of vegetation structure, commonly diversity and productivity. The independent variables are ecosystem or environmental parameters as well as grazing variables. Grazing variables are regulated by humans. Depending on different grazing regimes, the range of these variables extends from the unregulated (by humans) grazing of native species, to uncontrolled- or free-grazing of domestic ruminants, to over-grazing situations, and to optimized stocking rates for profitable animal mass gain per unit area of grazed land.

A meta-analysis of data on the effects of grazing on vegetation and soils, from more than 230 studied sites worldwide, has shown that the typical symptoms of the disturbances caused by domestic ruminants are changes in species composition, changes in dominant species, life-forms, and growth forms, changes in aboveground net primary production, and, finally, effects on the relationship between species, aboveground net primary production, root mass, and soil nutrients. In general, herbivory prohibits the most productive species from dominating and suppressing through competition the less productive ones. The conceptual model of Milchunas, Sala, and Lauenroth (1988) predicts the variation in plant diversity in relation to grazing intensity along gradients of moisture and evolutionary history of grazing (Table I). According to this model, grazing should have a greater effect on species composition in more humid areas because adaptations of tall growth forms capable of competing for light in a dense canopy are opposite to those that provide resistance to or avoidance of grazing. In contrast, plant adaptations to frequent loss of organs from drought or herbivory are similar,

TABLE I

The Plant Diversity–Grazing Intensity Relationship under Various Moisture and Grazing History Conditions, According to the Predictions of the Conceptual Model of Milchunas, Sala, and Lauenroth (1988)

		Moisture gradient	
		Semi arid	Subhumid
Evolutionary history of grazing	long	monotonic (linearly decreasing)	unimodal symmetric (parabolic)
	short	unimodal with steeply decreasing part	left skewed unimodal

and under arid and semiarid conditions, competition is primarily for belowground resources. The explanation of the evolutionary history of grazing gradient in the conceptual model is that increasing grazing history over evolutionary time results in greater capacities for regrowth following herbivory, and thus favors prostrate growth forms. In communities of long evolutionary history and high aboveground net primary productivity, grazing causes rapid shifts in the dynamic balance between suites of species adapted to either grazing avoidance/tolerance or competition in the canopy.

Field research offers evidence that herbivory by domestic ruminants may cause either an increase or a decrease in plant diversity. For example, in salt marsh habitats, heavy grazing eliminates sensitive species and produces a dense cover of graminoids in upper marsh coastal habitats. In other marshes, grazing produces bare patches that allow annuals and other low marsh species to invade upper marsh zonal communities. A retrogression in plant succession may occur in salt marshes and salt deserts because of heavy grazing. Intermediate levels of grazing by sheep, cattle, and horses could produce communities with the highest species richness and heterogeneity. At the other extreme of the aridity gradient, in a seasonally dry tropical savanna, species diversity was higher in a grazed area than in the neighboring area where grazers were permanently excluded. Furthermore, similarity in species composition between the grazed and ungrazed areas was very low.

The introduction of domesticated animals to a pasture or rangeland invariably causes changes in species composition, leading to an increase in the abundance of those species that are less palatable to animals. Low-growing, prostrate growth forms are selected by grazing, and annuals and shrubs appear to be favored. Tall perennials decrease in abundance because this growth form, which offers competitive advantages in dense canopies free from grazing conditions, exposes the plants to selective herbivory pressure in grazed ecosystems.

Between sites, aboveground primary productivity is negatively affected by grazing in the most productive ecosystems. Within-sites, lowlands are more affected than uplands. More productive lands are more likely to be preferred by grazing animals and breeders.

Grazing has negative impacts on root systems. Intensive grazing removes the product of vegetative growth, which in turn reduces root growth. Under drought conditions, shallow roots cannot fully exploit subsoil moisture reserves, and the plant becomes stressed.

B. Erosion, Desertification, and Land Degradation

Cattle are important agents of geomorphological change. The animals' impact on the landscape is to create bare soil by weakening the vegetation cover and then by breaking this cover down by trampling. On uplands, heavy grazing compacts the soil, reduces infiltration, increases runoff, and increases erosion and sediment yield. However, light and moderate grazing has effects that are much less significant. In riparian zones, grazing decreases erosional resistance by reducing vegetation and exposing more vulnerable substrate. Trampling directly erodes riverbanks, thus increasing turbulence and consequent erosion. Trampling also maintains and expands the area of bare soil upon which frost, rain, and wind act.

Livestock keeping when practiced in a nonsustainable way is one of the major causes of desertification. Desertification is defined as land degradation in arid, semiarid, and dry subhumid areas resulting from temporary climatic crises, especially droughts that occur periodically, and harmful human activities in vulnerable ecosystems. Land degradation leads to a reduced capacity of dryland areas to produce useful outputs—

crops, fodder for grazing livestock, bush and tree cover—or to sustain wildlife. Degradation of dryland areas involves a range of processes: the erosion of soils through water and wind, falling levels of soil fertility and damage to soil structure, loss of vegetation cover and change in species composition, reduced availability and decline in the quality of water supplies, loss of wildlife, and a decline in the biological diversity of plant and animal life. Such processes reduce the productivity of crops and livestock systems in dryland areas, and increase the vulnerability to food crisis of populations depending on these resources.

The processes of desertification were first recognized in the Sahelian region of West Africa. Overstocking and overgrazing, firewood collection, and cultivation of unsuitable soils are responsible for 80% or more of the desertized lands in Africa. However, dryland degradation is now accepted as a worldwide problem. Over the last decade, surveys have been carried out on a global level to assess the extent of desertification and soil degradation by region, and they have produced significantly differing results. The GLASOD survey (Global Assessment of Soil Degradation), commissioned by the United Nations Food and Agriculture Organization (FAO), showed that 19.5% of drylands worldwide were suffering from desertification. In contrast, the United Nations Environment Programme (UNEP) usually quotes an estimate of 70% of the world's dryland areas as suffering from some degree of desertification, with an estimated 900 million people worldwide at risk from problems of degradation. This survey includes not only areas affected by soil erosion, but also where a change in vegetation had occurred (e.g., where perennial grasses had been replaced by annuals). All regions of Africa have been affected by drought conditions and human pressures on land, as have parts of Mediterranean Europe, North America, Asia, and Latin America. In Spain alone, data from 1993 suggest that almost 1 million hectares (ha) of land are already considered as desert lands and another 7 million ha have been identified as being at high risk of desertification. In the United States, 90 million ha are considered to be affected by desertification.

The underlying factors that cause such adverse effects are many and various, and operate at different levels. At the local level, for example, inappropriate range, water, and livestock management practices may accelerate rates of erosion, especially on sloping land. Information continues to accumulate on the effects of digging deep boreholes in several regions of Africa. These boreholes discharge several liters of water per second, without the enforcement of any range management or land use policy, and have too often resulted in large concentrations of livestock (20,000 to 40,000 head) during the dry season. This destroys the range in the vicinity of the well over a radius of 20 km (125,000 ha) in one or two seasons (the stock rate being 10–15 times the carrying capacity).

At the national level, government policies on land tenure and use determine whether production processes will be sustainable or not. During the second half of the twentieth century, the Autonomous Russian Republic of Kalmykia has undergone severe desertification. Under Soviet rule, rangelands were increasingly devoted to animal production, and pastures were converted to cropland in a campaign to increase crop productivity. Pastures were grazed at rates that were two or three times their sustainable production, saiga antelope (*Saiga tatarica*) populations and habitat greatly decreased, more than 17 million ha were subjected to wind erosion, 380,000 ha were transformed into moving sands, and 106,000 ha were ruined by secondary salinization and waterlogging. By the 1990s, almost 80% of the Republic had undergone desertification, and 13% had been transformed into a true desert.

At the international level, rising prices provide encouragement to breeders and farmers to produce more and, at the same time, raise revenues that can be invested to increase capacity. The African arid zone harbors about 55% of that continent's 550 million head of livestock, and these livestock numbers have increased by 75% from 1950 to 1985, in spite of severe droughts that occurred in most arid zones in the early 1970s and 1980s. This is exponential growth of about 0.7% per year, compared to the 1.0–1.5% demographic growth rate of the African pastoralist population.

C. Nutrient Cycling and Over-Enrichment in Grazed Ecosystems

Nutrient cycles in a grazed ecosystem do not differ from those in an ungrazed one in terms of individual element pools, flows, potential input and output pathways, and interrelationships among various nutrient pools (Fig. 1). On the contrary, the sizes of the pools, the rates of biogeochemical processes, the flows, and the residence time in the various pools of elements and nutrients are greatly affected by herbivores and the breeding practice and stock size. For example, in natural pastures, animals remove aboveground plant biomass (shoots, stems, and leaves) and deposit excrement; in managed pastures, where grazing intensity exceeds the natural carrying capacity, additional herbage and other fodder is provided to the animals and, thus, additional inputs of

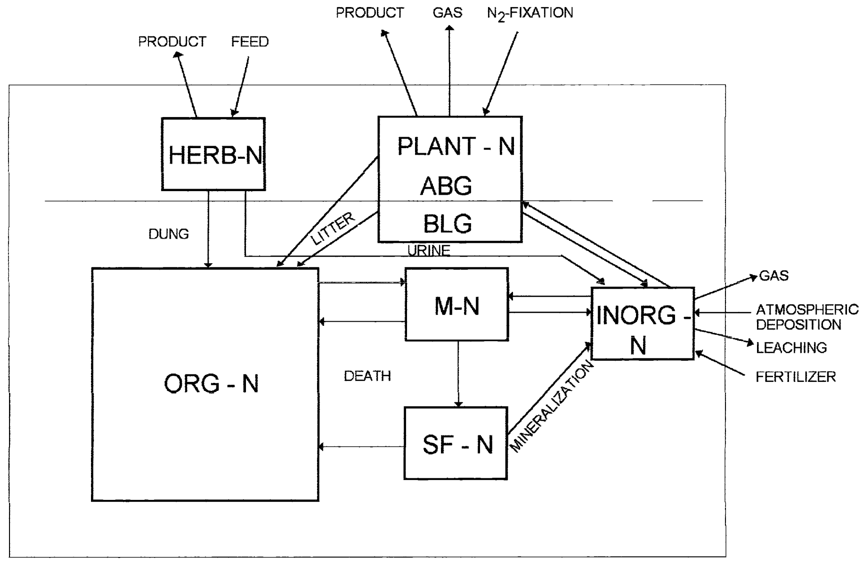


FIGURE 1 Pools and fluxes of nitrogen (N) in a typical pasture. Nitrogen pools are shown in their relative average sizes: ORG-N, organic soil N (about two orders of magnitude larger than other pools); M-N, microbial biomass N; SF-N, soil faunal biomass N; INORG-N, inorganic (NH_4^+, NO_2^-, and NO_3^-) soil N; PLANT-N, plant biomass N (ABG: shoots; BLG: roots); and HERB-N, large herbivore biomass N. The large rectangle represents field boundaries. For clarity, not all directions of flux are shown.

nutrients are entering the system. The cycle of nitrogen, most likely the most perplexing of all nutrients, is certainly disturbed under grazing conditions. In natural pastures, less litter returns to the soil, the topsoil organic matter declines, and the nutrient cycles are quantitatively disturbed. Grazers shorten the N cycle, increase the rate of N cycling, and cause significant redistribution of N, among pools and in space. On a daily basis, cattle excrete a somewhat constant amount of N in feces (8 lb/1000 lb of dry herbage consumed), the remainder being excreted in urine. Because they excrete more at night than during the day, bedding or watering places receive higher quantities of excreta-carried nutrients. Cattle and sheep retain a small fraction of the N consumed, and thus they have an impact on the control of N cycling that is out of proportion to their low biomass on an areal basis.

Areas of pasture that receive dung and urine may undergo marked changes in botanical composition. Urine spots and fecal pats contain the equivalent of 500 to 1000 lb N/acre, which is of different availability to plants: the N in dung is mainly in organic form, with a slow overall mineralization, whereas in urine N is present as urea, which is hydrolyzed rapidly to forms available to plants. The plant recovery of nutrients from excreta spots is rarely greater than 30%. Generally, the nitrogen in urine stimulates grass growth and the phosphorus in dung stimulates legume growth, especially on P-deficient soils. However, plants immediately beneath dung pats may be killed and urine occasionally scorches them.

The cycles of nitrogen, phosphorus, and other nutrients are not generally closed at the field or farm level. For nitrogen, six main pathways of loss have been described in grazed ecosystems: NH_3 volatilization, denitrification, wind and water erosion, NO_3^- leaching, fire, and incorporation into animal biomass (exported or retained *in situ*). The relative importance of these pathways depends on environmental conditions and breeding practices. In humid and subhumid ecosystems, N losses are related to leaching, whereas volatilization is most important in semiarid regions, and the export of animal products and erosion in deserts.

Runoff from agricultural land and livestock feedlots is among the major "nonpoint" sources of nutrients entering marine and freshwater ecosystems, causing pollution and eutrophication. This has been shown on very large scales, as in the United States, where these widely dispersed activities are the major source of water pollution, as well as at the level of local ecosystems, such as lakes, rivers, and estuaries. In aquatic ecosystems, overenrichment with P and N causes a wide range of problems, including toxic algal blooms, loss of oxygen, fish kills, loss of seagrass beds and other submerged aquatic vegetation, degradation of coral reefs, and loss of biodiversity—including species important to commercial and sport fisheries and shellfish industries.

Lake studies allow contemporary sediment and nutrient dynamics to be placed in a historical context so that trends and rates of change in catchment inputs may be calculated. Surface runoff from heavily grazed grassland has a high suspended sediment, ammonium nitrogen, and particulate phosphorus load. The combined effect of the long-term increase in the organic loading from livestock and the inorganic N and P load from fertilizers may be the source of nutrient enrichment in lakes. The literature offers several examples of eutrophication studies in lakes and streams; for example, in the already extinct karstic Jastericie Lake in Slovakia, in the coastal freshwater systems of Slapton Ley and Loe Pool in southwestern England, and in the streams draining the Mount Lofty ranges in South Australia, livestock breeding is the main source of organic N and P that affects the trophic status of surface waters.

Intensive animal production generally involves feeding large numbers of animals in small areas. For example, 4% of the cattle feedlots in the United States produce 84% of the cattle. Nutrients in manure can be recycled by applying the manure to cropland. However, the amount of manure generated by concentrated livestock operations often far exceeds the capacity of nearby croplands to use and retain the nutrients. Thus, excess fertilization and manure production cause a surplus of N and P, which accumulate in the soil. Some of this surplus is transported in soil runoff to aquatic ecosystems. Especially for N, the surplus is mobile in many soils, and much leaches into surface waters or percolates into groundwater. Indeed, the density of animals on land is directly related to nutrient outflow to aquatic ecosystems. Surplus N can also volatilize in nitrous (NO_2) and ammonia (NH_3) forms to the atmosphere. The emission of ammonia from stored and land-applied manure can result in a significant loss of nitrogen for crop production. A high atmospheric concentration of ammonia can result in the acidification of land and water surfaces, cause plant damage, and reduce plant biodiversity in natural systems.

Nitrous oxide, together with carbon dioxide and methane, are considered to be major greenhouse gases. Besides emissions of these gases from the plant–soil compartment of the pasture soil system, CO_2 and CH_4 emissions from cattle and their excreta are significant components of the global fluxes. Methane emitted from dung is a very important contributor to the global methane budget; the corresponding figure is as high as 20%. Grazing animals on managed pastures and rangelands have also been identified recently as significant contributors to the global N_2O budget. This occurs because of the concentration of herbage N in urine and dung patches, and by the compaction of the soil due to treading and trampling. The limited amount of experimental data indicates that 0.1 to 0.7% of the N in dung and 0.1 to 3.8% of the N in urine is emitted to the atmosphere as N_2O. Integral effects of grazing animals have been obtained by comparing grazed pastures with mown-only grassland. Grazing-derived emissions, expressed as a percentage of the amount of N excreted by grazing animals in dung and urine, range from 0.2 to 9.9% with an overall mean of 2%. Using the emission factor and data statistics from FAO for numbers of animals, the global contribution of grazing animals is estimated at 1.55 Tg N_2O-N per year.

III. BIOTIC RELATIONSHIPS OF LARGE DOMESTIC ANIMALS

A. Reactions of Plants to Grazing

The direct act of grazing represents a loss of organs or parts of organs to individual plants and an alteration of canopy structure to the community. Resistance of plants to grazing involves avoidance and/or tolerance mechanisms.

Plant tolerance reactions to the grazing event will depend on the capacity of individual plants to compensate for lost organs and the relative impact of the removal on competitive relationships in the canopy. Various species of animals graze differently because of their prehension organ anatomy and method, behavior, and diet preferences. For instance, cattle graze individual plant parts (leaves or stems) less selectively than do sheep and goats. The literature offers examples of strong interactions among intensity, severity, and other properties of grazing of different animal species and pasture (species) in terms of various components of plant fitness. Especially for grass species, the height, tiller number, survival, and reproduction are the fitness components that are most usually monitored in grazed plots/individuals and ungrazed controls.

Three alternative hypotheses on the effects of grazing intensity on plant growth and fitness have been proposed (Fig. 2). The first hypothesis predicts that net primary productivity shows a consistent decline as the intensity of grazing increases. The second hypothesis

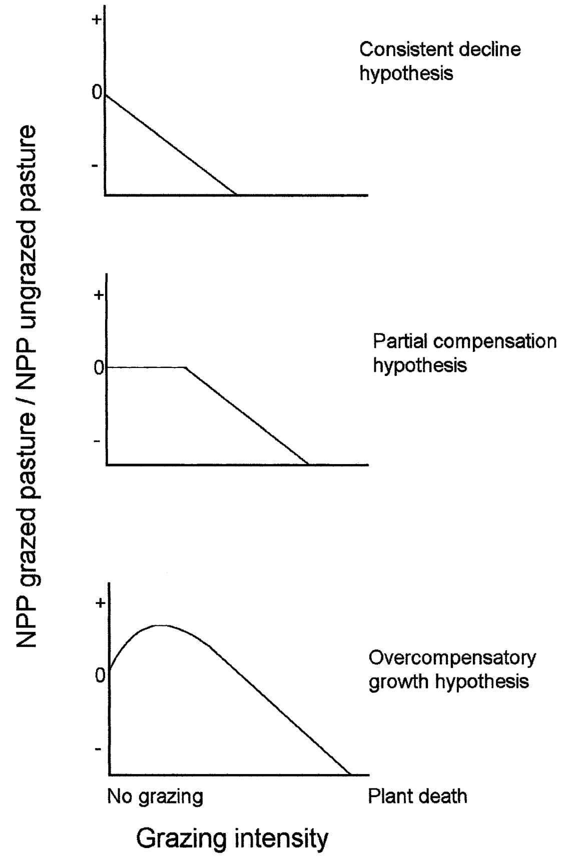


FIGURE 2 Theoretical relationships between the variation of net primary productivity (NPP) of grazed and ungrazed grassland pastures and the grazing intensity. Lines represent possible trends in NPP as grazing intensity increases, according to the three examined hypotheses: consistent decline hypothesis, partial compensation hypothesis, and overcompensatory growth hypothesis.

predicts that plants compensate for tissue removal up to some level, beyond which productivity begins to decline. The third hypothesis, known as "overcompensatory growth," states that within some levels of herbivorous feeding, plant productivity may be enhanced before declining beyond a threshold of grazing intensity. This hypothesis has generated an ongoing debate on the controversy of plant response to grazing, such as herbivore optimization and overcompensation.

Observations of mixed cattle and American elk grazing in high-elevation rangeland conditions in southwestern North America show that vegetation has deteriorated. Experiments in the North American tallgrass prairie suggest that overcompensation is a nonequilibrium plant response to grazing. Photosynthate that would be stored as reserves and used for root growth and flower and seed production is instead used to replace lost leaf area, thereby resulting in higher foliage productivity. However, under chronic grazing or mowing, vegetation is prevented from maintaining high nutrient and water uptake capacity (large root biomass) and accumulating reserves that allow overcompensation responses.

Herbivory by large animals is known to function as a selection pressure to increase herbivory resistance within plant populations by decreasing the frequency of genotypes possessing large erect canopies. Data on the trade-off between herbivory resistance and competitive ability in *Schizachyrium scoparium* confirm that herbivory by domestic cattle may function as a selection pressure to induce architectural variation in grass populations within an ecological time frame (ca. 25 years).

Comparative studies between species suggest that there is no consistent pattern of grazing effects on survival, reproduction, recruitment, and regeneration. In one study, five abundant native Kansas tallgrass prairie perennial forbs (*Baptisia bracteata, Oenothera speciosa, Vernonia baldwinii, Solidago missouriensis*, and *Salvia azurea*) were chosen to examine the effects of native (bison) and domestic (cattle) ungulates on plant growth and reproduction. The results show that their responses are complex and vary significantly among plant species, ungulate species, and plant life-history stages. *Baptisia bracteata, O. speciosa*, and *V. baldwinii* increased in growth and reproduction in grazed sites, indicating competitive release in response to selective grazing of the dominant warm-season matrix grasses. Species with reduced performance in grazed sites are likely to be negatively affected by disturbances generated by ungulate nongrazing activities, because none of the forbs studied was directly consumed by bison or cattle. Furthermore, the native and domestic ungulates differ significantly in their effects on forb growth and reproduction.

Among 18 species of shrubs and trees in southeastern Australia, 10 showed significant negative effects on recruitment and/or regeneration from present or past sheep grazing. In this case it was shown that the negative effect of sheep grazing on recruited seedlings must exceed that of natural thinning before overall regeneration is affected. Evidence that grazing by sheep and goat increases the total inclusive fitness in the Mediterranean shrub *Anthyllis cytisoides* has been reported: in this species, moderate grazing promotes growth, stability of vegetative structures, and adult survival, and a drop in seed production. Direct consumption of reproductive organs has significant effects on the overall fitness of the grazed species. In the annual wild wheat *Triticum* *dicoccoides*, removal of maturing inflorescences by cattle reduces its fitness by 50%, estimated as the number of seeds produced per seedling. In this species, protection from grazing leads to a significant increase in the number of individuals that produce mature inflorescences as well as in the number of spikelets per inflorescence. In the clonal species *Yucca elata*, cattle browsing of inflorescences may reduce reproductive effort, but the most significant changes in population structure are due to the browsing of small caudices, including both genets and ramets. Thus, the expansion and dominance of "resistant" species on heavily grazed land may be a result not only of reduced competition, as is commonly assumed, but also of enhanced performance of some of these species.

Grazing avoidance mechanisms reduce the likelihood of defoliation by animals. For example, leaf surface chemistry of plants is related to the extent of defoliation by livestock. Data on cattle, sheep, and goat herbivory on tarbush (*Flourensia cernua*), an abundant but generally unpalatable Chihuahuan Desert shrub, support the hypothesis that secondary chemicals—mostly terpenes and phenolic compounds—in its resinous leaves may influence the diet selected by the ruminants. Although individual leaf surface compounds do not appear to greatly affect the degree of use of the plant by livestock, collectively these compounds may partially explain the differential herbivory on tarbush plants by livestock. Mimicry, both chemical and morphological, has also been suggested as a grazing avoidance mechanism for plants. Experimental tests with sheep showed that odor alone is not persistently effective in preventing herbivory, but that both taste and odor must be similar for one plant to successfully mimic another.

Spines and thorns are considered to be defense structures against herbivory by both wild and domestic animals. Experimental evidence from *Acacia drepanolobium* suggests that spine length is an inducible defense, with longer spines being produced by branches experiencing a greater level of herbivory. Examination of *Acacia* trees protected from herbivory for several years suggests that reduction in spine length eventually exceeds 70%. The effectiveness of spines and thorns as an antiherbivore defense in several plants in arid Australia did not vary with the evolutionary history of the herbivores (i.e., wild versus domestic). Furthermore, additional evidence shows that the interaction of ants of the genus *Crematogaster* and thorns of *A. drepanolobium* is a means of defense against browsing goats. This interaction causes the animal to stop feeding almost immediately, therefore keeping the amount of foliage lost to a minimum. It is hypothesized that the acacia–ant relationship evolved partly because of the pressure from browsing herbivores.

B. Seed Dispersal and Germination

Throughout the world, the presence of wild or domestic ruminant herbivores is correlated with the maintenance of high levels of plant diversity in natural and seminatural grazed systems, where the grazing pressure does not exceed the carrying capacity of the vegetation. Zoochorous dispersal by herbivorous mammals has been verified repeatedly and its possible influence on the structure, function, and diversity of plant communities, mainly herbaceous, has been suggested.

Cattle, sheep, and goats are important vectors for endo- and epi-zoochorous seed dispersal. On large scales, seed dispersal systems associated with domestic ruminants have been proved to be particularly favorable for the introduction of alien plant species, and herbivores have facilitated the naturalization and spread of many alien herbaceous species from their initial points of introduction. The case of central Iberian herbaceous species introduced into the Mediterranean-type zone of Chile is very relevant: almost 15% of the central Iberian herbaceous species are naturalized in Chile, with the endo- and epi-zoochorous species representing 21 and 23% of them, respectively. The expansion of the distribution of individual species is often correlated with the effects of cattle ingestion on the viability and germination rates of seeds. For example, the calden (*Prosopis caldenia* Burkart) is the dominant tree of the xerophytic open forest in the semi-arid pampa of Argentina. Calden has gradually increased its distribution throughout the region during the past century as a result of livestock grazing in the pampa forest. Excreted calden seeds display a range of delayed germination responses. This variation would increase the probability of seed germination for a variety of environmental and site conditions.

The potential dispersal distances for adhesive seeds attached to the fur of cattle range from tens of meters to kilometers. The morphology of the seed's structure and its position on the animal's body influences the length of time that it remains attached to the fur.

Endozoochorous dispersal selects for traits that enhance ingestion and passage of viable seeds through the animal. Passage of buffalo grass (*Buchloe dactyloides*) seeds, one of the two dominant grasses of the North American shortgrass prairie, through cattle has a positive effect on germination and seedling growth from intact diaspores; the damage due to mastication is minimal, and the retention time is from 1 to 5 days. This combination of retention time and movements of the animals influences the spatial expansion, distribution, and abundance of this species. Clear experimental evidence that the germinability of seeds increases significantly following their passage through the cattle gut has been provided for the legume *Biserrula pelecinus,* which is greatly favored by its dispersal through cattle dung.

The amounts and diversity of viable seeds contained in the feces of domestic ruminants grazing in nature may be quite high. Based on germination trials, the number of seed species germinated in the dung of feral cattle feeding in savannas and floodplain wetlands in Rajasthan, India, was about 450 seeds per m^2, belonging to 35 species. Similar results have been obtained in Mediterranean open woodlands, where cow dung may yield as much as 70 seeds per gram of dry dung from a large number of species (about 75 species). Although ruminant species differ in their traits and feeding habits, dispersal seems to be mainly determined by seed production of the plant community.

Large isolated trees are a common feature of the agricultural landscape in humid tropical regions that were originally covered by rain forest. These isolated trees are used as a source of shade for cattle, but they also function as nursery plants for rain forest species by facilitating the establishment of zoochorous species whose seeds are deposited under their canopies by cattle, frugivorous birds, and bats. The same pattern is observed in Botswanan savannas: the analysis of seed pools under the canopy trees *Dichrostachys cinerea* showed that epizoochorous species such as *Tragus berteronianus* were dominant at the cattle resting sites under trees.

Patch dynamics are also affected by seed zoochorous dispersal. In Mediterranean pastures, cattle-dispersed endozoochorous seeds germinate in manure and colonize the dung patches. The micro-succession involved is independent of the type of pasture. A small-scale spatial pattern results in which gaps of old dung are dominated by endozoochorous species. Thus, dung patches enhance the similarity between different communities when they are grazed, but also increase the variation within communities. In Australian sub-alpine grassland vegetation, the density of shrub seedlings in gaps varied considerably in space, but it is related to the dispersal of seeds and the trampling and browsing effects of domestic cattle.

Large seeds such as *Quercus suber* acorns are rapidly predated by large herbivores. However, in this case, experimental data seem to suggest that seedling emergence rate is inversely related to the intensity of predation on the acorn bank.

C. Domestic Animals as a Food Source for Wild Predators

Domestic ruminants have largely replaced wild grazing animals over large areas on all five continents. Natural predators and scavengers of large mammals have easily adapted to this alternative food source. Population densities of cattle, sheep, and goats are equivalent and often higher than the densities of the wild grazers that have been replaced, as the animal keepers' tendency is to maximize herd or flock size and food may be supplemented at times of natural fodder shortage. Furthermore, domestication has usually produced heavier, slower-moving, less alert, and less experienced animals, which often have their ability to move impaired by various means and structures such as pens, fences, and tethering. Thus, domesticated animals are generally speaking easy prey for most natural predators. The way of life of pastoral societies around the world has been profoundly influenced by the constant strive for protection of their flocks and herds from natural predators.

Wolves (*Canis lupus*) are legendary in Western culture as predators of cattle, goats, and especially sheep. Today they have been exterminated or reduced to small populations over much of their former range in the Northern Hemisphere, but where they do exist they are still important predators of livestock. In regions like parts of southern Europe, where stock-raising is prevalent and wild prey are rare, their presence is still connected with the raising of stock animals. The wolves either prey directly on live individuals, most often stray ones, or feed on carcasses and offal. Similarly, dingos (*Canis familiaris*) can be serious predators of sheep in Australia, as are coyotes (*Canis latrans*) in North America. In Africa, Asia, and North and South America, large cats such as leopards (*Panthera pardus*), tigers (*Panthera tigris*), and pumas (*Puma concolor*) commonly prey on cattle, sheep, and goats. Big cats and the other canids differ from wolves in that their existence does not seem to rely to any significant extent on domestic animals. The big cats are especially dependent on extensive tracts of undisturbed habitat where natural prey is available, and consequently domestic animals are consumed opportunistically and usually only by certain individuals.

In many parts of Eurasia, both golden eagles (*Aquila chrysaetos*) and lammergeiers (*Gypaetus barbatus,* German for "lamb vulture") have traditionally been regarded as significant predators of lambs, and this has been the reason for their local extermination. In reality they kill very few lambs but commonly feed on dead lambs or the carcasses of older sheep. Both Old and New World vultures also include dead cattle, sheep, and goats in their diet. As in the case of the wolf, in parts of the world where wild ungulates are now rare, such as around the Mediterranean, these sources of food are essential for the survival of vulture populations. Although griffon vultures (*Gyps fulvus*) are not migratory, they perform movements between summer and winter quarters, closely following flocks of sheep as they move from high-altitude summer pastures to lowland grazing land in winter. For the larger eagles and vultures, open range grazing has a dual functionality in providing a food source (essential or supplemental) and by indirectly maintaining open habitats where such species are able to forage.

Cattle, sheep, and goats appear in the diet of the majority of other medium or large natural predators such as bears (*Ursus* spp.), foxes (*Vulpes* spp.), jackals (*Canis* spp.), other eagles (*Aquila* spp.), and buzzards (hawks, *Buteo* spp.) to a varying extent, depending on the region, the species of predator, and the method of stock-keeping. In most such cases, carcasses, offal, and still borns are consumed rather than healthy, full-grown adults.

D. Domestic Animals and Invertebrate Fauna

Both positive and negative effects of cattle, sheep, and goats on invertebrate communities of grazed ecosystems have been reported. In grazing systems where the carrying capacity of vegetation is not exceeded by grazing, as in the traditionally managed European mountain meadows or mixed woodlands, the maintenance of a landscape mosaics favors the persistence of high plant diversity and abundance of flowers and has positive consequences on insect population dynamics. Bumblebees, butterflies, syrphids, and other insect groups foraging on flowers depend on the landscape management in which domestic grazers play an important role. In central Spain, traditional landscape management for non-intensive grazing by goats and cattle favors the conservation status of the butterfly *Euphydryas aurinia,* which depends on patches of open oak woodlands mixed with open areas, where important nectar sources and larval foodplants are present. However, indirect effects on pollinators may appear from the removal of flowers by grazing: in *Yucca elata,* cattle preferentially consume inflorescences, which are found to be highly nutritious. This implies the possible local extinction of the yucca moth *Tegeticula yuccasella,* the exclusive pollinator of the plant.

In the shortgrass prairie of Colorado, in the United States, heavy grazing negatively affects the colony density and distribution of harvester ants (*Pogonomyrmex occidentalis*) on a broad scale, with important consequences to behavioral, community, and ecosystem processes. Browsing by goats has been proven to increase the damages from both tunneling (Diptera) and “blotch-making” (Lepidoptera) leaf miners on trees growing on the Aldabra Atoll.

Insects are particularly sensitive indicator taxa of land use (especially the Cicindellidae, Staphylinidae, and Carabidae). Domestic cattle influence the arthropod diversity through trampling, its intensity being more important than the type of trampling. In semi-arid tropical soils in Queensland, Australia, heavy grazing significantly affects Acari populations as well as the diversity and activity of termite species. The deterioration of soil hydraulic properties associated with cattle trampling at high stock rates is responsible for these negative effects on detritivorous termite activity in the topsoil.

A wide range of invertebrates, as well as bacteria and fungi, are involved in the breakdown and eventual decomposition of dung. Some are facultative generalist decomposers of organic matter, but there are also a large number of specialist dung consumers, such as dungflies (Scathophagidae) and dung beetles (Geotrupidae, Scarabaedae). The array of dung-feeding invertebrates present in a given area is mainly dependent on the characteristics of the dung, that is, on the taxonomy of the dung producer and its diet rather than on a distinction between wild and domestic stock.

E. Domestic Animals and Birds and Small Mammals

Excluding predators and scavengers, the main influence of cattle, sheep, and goats on other animals occurs indirectly through the creation and maintenance of open pasture land, in particular in areas where the natural vegetation cover would be forest or scrub. Tall-grass meadows, short turf on stabilized sand dunes, alpine meadows, dry stony pastures, and all other forms of grazing land that exist on different soils and under different climatic regimes are all attractive to species that would otherwise be restricted to steppe, desert, or arctic habitats on a global scale or to the margins of wetlands, beaches, steep hill slopes, and woodland clearings on a more local scale. The creation of pastures has gone hand in hand with the creation of agricultural land, which has also provided vast areas of open habitat. The expansion of both pasturelands and croplands began several thousand years ago, and the process of colonization and spread of open-country species into such newly created habitats is lost in prehistory and is largely a matter of conjecture.

Animal species closely connected with grazing land include mammals such as moles (*Talpa* spp.) and voles (Microtinae), birds such as larks (Alaudidae) and plovers (Charadriidae), and insects such as grasshoppers (Acrididae) and ants (Formicoidea). Grazing is often employed as a management tool, for example, to create the very short turf required by wintering white-fronted geese (*Anser albifrons*) in Britain.

Other species may use cattle, sheep, or goats more directly in feeding. Two well-known cases are cattle egrets (*Bubulcus ibis*) and yellow wagtails (*Motacilla flava*). Cattle egrets commonly use cattle, as well as other domestic and wild ungulates, as a vantage point and vehicle, capturing large insects disturbed by the cattle as they move. Yellow wagtails may also do the same or they may feed on flies and other insects attracted to the animals themselves or their dung.

Negative impacts on other animal species may also occur. For example, when natural grazing land is used, domestic ruminants may compete with wild herbivores. Although wild herbivores may suffer, human interest is usually focused on the reverse impact of wild herbivores on domestic ones [e.g., rabbits (*Oryctolagus cuniculus*) and sheep in Australia].

Another negative impact on wildlife is the trampling of eggs and young of ground-nesting birds. This problem is more frequent in wetland habitats, where most bird species nest on the ground, and may be particularly damaging for colonially breeding species such as terns (Sternidae) and avocets (*Recurvirostra* spp.). Although these birds generally select islets or other sites that are inaccessible to ground predators (e.g., foxes) and thus also inaccessible to livestock, problems may be caused when water levels change or when a herd or flock is intentionally transferred to such a site.

F. Domestic Ruminants as Hosts for Parasites and Pathogens

Domestic ruminants harbor significant numbers of internal and external parasites. Helminths, protozoa, and arthropods (especially ticks and mites) are the most common components of the parasitic communities, although many other species of occasional or temporary parasites, such as blood-sucking Diptera, also infect domesticated animals. Endoparasites, mainly gastrointestinal parasites, have received most attention in veterinary research because of their prevalence as disease-causing organisms that are economically important in rearing commercial livestock. Within a species, the taxonomic diversity of parasites varies between the various organs of the hosting organism. Thus, the digestive tract, the liver, the circulatory system, the respiratory tract, the skin and the subcutaneous tissue, the muscles and the tendons, the eyes, the central nervous system, and the serous cavities are affected by quite distinct species of parasites such as trematodes, cestodes, nematodes, protozoa, and arthropods. For example, in sheep and goats, the list of gastrointestinal parasites comprises 15 trematodes, 9 cestodes, 88 nematodes, and 25 protozoa; the parasitic fauna in the skin and the subcutaneous tissues comprises only 6 nematodes but 78 arthropods. Parasitic communities also differ between host animal species, even closely related races. Likely causes are differences in host suitability and feeding differences that affect the probability of transmission.

Besides the specificity of the host–parasite interactions, the diversity of the domestic ruminants' parasite communities is strongly influenced by species–area and species–climate relationships. As an example, for nematodes it has been demonstrated that

(a) large areas of permanent pasture include a variety of microenvironments that are favorable to the development of the free-living stages of various species of nematodes and
(b) free-living stages are very susceptible to dryness and survive better in areas with heavy rainfall.

The widespread use of veterinary chemicals against endoparasites, pests, and other pathogens of domestic animals characterizes the high-input livestock breeding and production strategy. Avermectins, a relatively new class of broad-spectrum pesticides, are used widely to control livestock parasites. Following veterinary treatment, avermectins are eliminated in the livestock feces. The dung mesofauna potentially exposed to avermectin (or other active compound) residues includes insects, earthworms, springtails, mites, and nematodes. The effects range from acute toxicity in larvae and adults, through disruption of metamorphosis, to interference with reproduction. For example, at high drug concentration in the dung, larvae of the dipteran Cyclorrhapha are killed or paralyzed, while at lower levels their metamorphosis is inhibited. At very low concentrations of avermectins, well below levels occurring in feces after routine treatment, adult emergence is reduced and a significant number of imagines show morphological abnormalities. Nematoceran Diptera are less sensitive than Cyclorrhapha, but larval and pupal development are affected at higher dose levels.

Dung mesofauna occupy a variety of different niches within the ecosystem and the faunal composition changes as the pats age. Some members of this fauna act in concert with soil microbial flora and assist in the breakdown of dung and consequent nutrient recycling on pasture-lands. A retardation in the rate of loss of biomass of dung pats from avermectin-treated cattle has been observed following the various forms of drug administration. Rare insects could be put at risk by the use of avermectins, especially those that breed exclusively in the dung of the herbivores on which avermectins are used.

The use of avermectins may also indirectly affect some species of vertebrates by depleting the quality and quantity of important food resources. The effects of any reduction in invertebrate food in livestock dung would be expected to be especially severe if it occurred at critical times for the vertebrates, such as during the breeding season or when newly independent young animals were foraging and fending for themselves. Insects that develop in livestock dung therefore have important, additional roles in the ecology of pasturelands other than aiding dung degradation processes. Livestock dung is an important feeding habitat for a number of vertebrate species. The potential for direct poisoning of vertebrates through the accumulation of avermectins following consumption of invertebrates containing residues would, on present knowledge, appear to be limited, but it should not be disregarded.

This issue has generated much controversy regarding the extent of the effects on pasture and rangeland biodiversity and ecosystem functioning. It has been suggested that ecotoxicological studies commonly disregard the veterinary use patterns of drugs and consequently they overestimate their negative effects at a populational level. Avermectins administration patterns in temperate regions indicate that peak periods of insect activity and peak times of avermectin use are often asynchronous. When avermectin usage and insect activity do coincide, the heterogeneous patterns of administration to livestock and the focus of treatment on young animals result in the deposition of feces that are predominantly free of avermectin residues. Results of large-scale, long-term studies indicate that, even under conditions of relatively high levels of avermectin use in cattle, the impact on non-target insect populations and their function is limited.

See Also the Following Articles

AGRICULTURE, TRADITIONAL • DESERTIFICATION • GRAZING, EFFECTS OF • GREENHOUSE EFFECT • RANGE ECOLOGY

Bibliography

Clarholm, M., and Berstrom, D. (eds.). (1989). *Ecology of Arid Lands: Perspectives and Challenges*. Kluwer Academic Publishers, Dordrecht, Netherlands.

Evans, R. (1998). The erosional impacts of grazing animals. *Progr. Phys. Geogr.* **22**(2), 251–268.

FORUM. (1993). Grazing theory and rangeland management. *Ecol. Appl.* **3**(1), 1–38. (Papers by S. A. Levin; E. L. Painter and A. J. Belsky; M. I. Dyer, C. L. Turner, and T. R. Seastedt; S. J. McNaughton; M. J. Trlica, and L. R. Rittenhouse; D. D. Briske; J. W. Bartolome; D. L. DeAngelis and M. A. Huston; I. Noy-Meir; D. T. Patten; and P. G. Risser.)

Milchunas, D. G., and Lauenroth, W. K. (1993). Quantitative effects of grazing on vegetation and soils over a global range of environments. *Ecol. Monogr.* **63**(4), 327–366.

Milchunas, D. G., Sala, D. E., and Lauenroth, W. K. (1988). A generalised model of the effects of grazing by large herbivores on grassland community structure. *Am. Nat.* **132**(1), 87–106.

Pain, D. J., and Pienkowski, M. W. (eds.). (1996). *Farming and Birds in Europe*. Academic Press, San Diego.

Popay, I., and Field, R. (1996). Grazing animals as weed control agents. *Weed Technol.* **10**, 217–231.

Russelle, M. P. (1992). Nitrogen cycling in pasture and range. *J. Production Agric.* **5**, 13–23.

Snaydon, R. W. (ed.). (1987). *Managed Grasslands: Analytical Studies. Ecosystems of the World, Vol. 17B*. Elsevier, Amsterdam.

Spratt, D. M. (1997). Endoparasite control strategies: Implications for biodiversity of native fauna. *Int. J. Parasitol.* **27**(2), 173–180.

Stuth, J. W., and Lyons, B. G. (eds.). (1993). *Decision Support Systems for the Management of Grazing Lands*, Man and the Biosphere Series, Vol. 11. UNESCO/The Parthenon Publishing Group, Rome.

CENTRAL AMERICA, ECOSYSTEMS OF

Rodolfo Dirzo
Universidad Nacional Autónoma de México

GLOSSARY

biogeography The study of the distribution of plant and animal life in the earth's environment and of the biological and historical factors that produced this distribution.

geomorphology The study of the surface configuration of the earth, especially the nature and evolution of current land forms, their relationships to underlying structures, and the history of geological activity as represented by such surface features.

land use change The changes in the way land is used at a given location, for example when a forested land is converted to agricultural fields or when a forestry plantation is changed to other types of use.

páramo This term refers to a cold, inhospitable and humid landscape located above the tree line.

plate tectonics A modern geological theory of tectonic activity according to which the earth's crust is divided into a small number of large, rigid plates whose independent movements relative to one another cause deformation, volcanism, and seismic activity along their margins.

CENTRAL AMERICA is a relatively small area extending from the narrowest part of southern Mexico to the Panama–Colombia border. Despite its small area and restricted latitudinal extension, Central America contains a significant proportion of the biological diversity of the Earth. This is due to its complex geological history, current diversity of climates, and topographic heterogeneity. This is reflected in the diversity of ecosystems present in this region, which in turn contain a remarkable diversity of species of plants and animals of tropical and temperate affinity, as well as many endemic organisms. The ecosystems of Central America include tropical rain forests, seasonally dry tropical forests, tropical cloud forests, temperate forests, and other high-elevation ecosystems, particularly páramo and high-elevation grasslands. Such concentration of biodiversity in Central America is threatened by global environmental changes, particularly the current patterns of land-use leading to massive deforestation and habitat fragmentation. In addition, climatic changes are likely to affect the distribution, functioning, and biodiversity of the ecosystems of the region, particularly the temperate and high-elevation ecosystems.

I. CENTRAL AMERICA AS A BIOGEOGRAPHIC REGION

The definition of Central America varies considerably among sources. The most popular notion of what constitutes Central America is based on a geopolitical criterion. According to this, Central America is that region of the Western Hemisphere that constitutes the isthmus lying between Mexico and South America, encompassing the countries of Guatemala, Belize, Honduras, El Salvador, Costa Rica, and Panama. Some geographers also include at least one (Jamaica) or most (Cuba, Haiti, the Dominican Republic, and Barbados) of the Antillean Islands as part of Central America, but the most accepted views only consider the continental part of the region as Central America. However, even this widely accepted geopolitical definition is not entirely satisfactory for the purposes of this encyclopedia. A description of the natural ecosystems of Central America requires a more natural definition. Thus, this article will consider a biogeographical and geomorphological approach, according to which Central America is more naturally defined as an elongated, tapering isthmus that begins in the narrowest part of southern Mexico, at the Isthmus of Tehuantepec (Fig. 1). This narrow region divides the area of volcanic rocks to the northwest from the folded and considerably faulted structures of the more conventional Central America to the east. The southernmost limit of Central America is the valley of Río Sucio (the Astrato River) in Colombia, located just to the east of the Panama–Colombia border, where the massive Darien jungles begin.

An important biogeographic aspect derived from the location of the area is that this narrow strip of land currently connects the neotropics of South America and the Nearctic zone of North America, two major biogeographic realms of Earth. An important additional implication is that Central America currently operates as a corridor for terrestrial organisms from both realms and as a barrier to marine organisms from the Caribbean and the Pacific seas. However, such situation has not always been the case. A variety of hypotheses have been suggested regarding the possible configuration of the area in the past. However, a consensus exists that from the latter Mesozoic, until approximately 5 or 6 million years ago, in the Pleiocene, no continuous terrestrial connection existed between South and North America via Central America. Moreover, throughout the largest part of the Cenozoic era (65 to approximately 5 million years ago), the region comprised from Nicaragua to the northernmost part of Colombia was probably an archipelago similar to the Lesser Antilles of today, providing only, perhaps, an occasional pass for terrestrial life forms. Most likely, the continuous connection that Central America currently provides between the north and the south has been in existence for only approximately the past 3 million years, and the mountainous backbone of the region did not reach its present elevation until the latter part of the Cenozoic. This means that the variety of habitats that currently characterize the region is a rather recent phenomenon and that the ecosystems and biodiversity of Central America in general are a recent blend that has resulted from a complex palebiogeographic history. A review of the paleobiogeographic history of the region is beyond the scope of this article; however, some salient aspects that, in addition to what I described previously, impinge on the current biodiversity of Central America include the following (Rich and Rich, 1983). Until approximately 3 million years ago the mammalian faunas of Central American were composed predominantly of North American elements and only after this time did a north–south interchange begin with a significant incursion of vertebrates from South America. It is possible that an endemic small-mammal fauna might have developed at about this time in Central America and southern North America and dispersed toward the south. The greatest affinity of Central American angiosperm (flowering plants) floras with South America indicates that dispersal from the south was more pronounced for plants than for animals, although Pleistocene cooling periods promoted dispersal of typically montane plants toward the south. In addition, at the height of the North American glaciation period in the Pleistocene, climatic conditions favored the dispersal of grassland- and savanna-adapted forms. Nevertheless, these ecosystems must have become as fragmented as they are today by the increasing precipitation of the latter part of this epoch. Finally, it was not until these times, toward the end of the Pleistocene, that the tropical rain forest ecosystem became widespread in Central America. This has the implications that this ecosystem, geologically speaking, is very young in the region and that its expansion decreased the effectiveness of the terrestrial connection between North and Central America. This has played a determining role in the modern composition of the biodiversity and ecosystems currently seen in Central America.

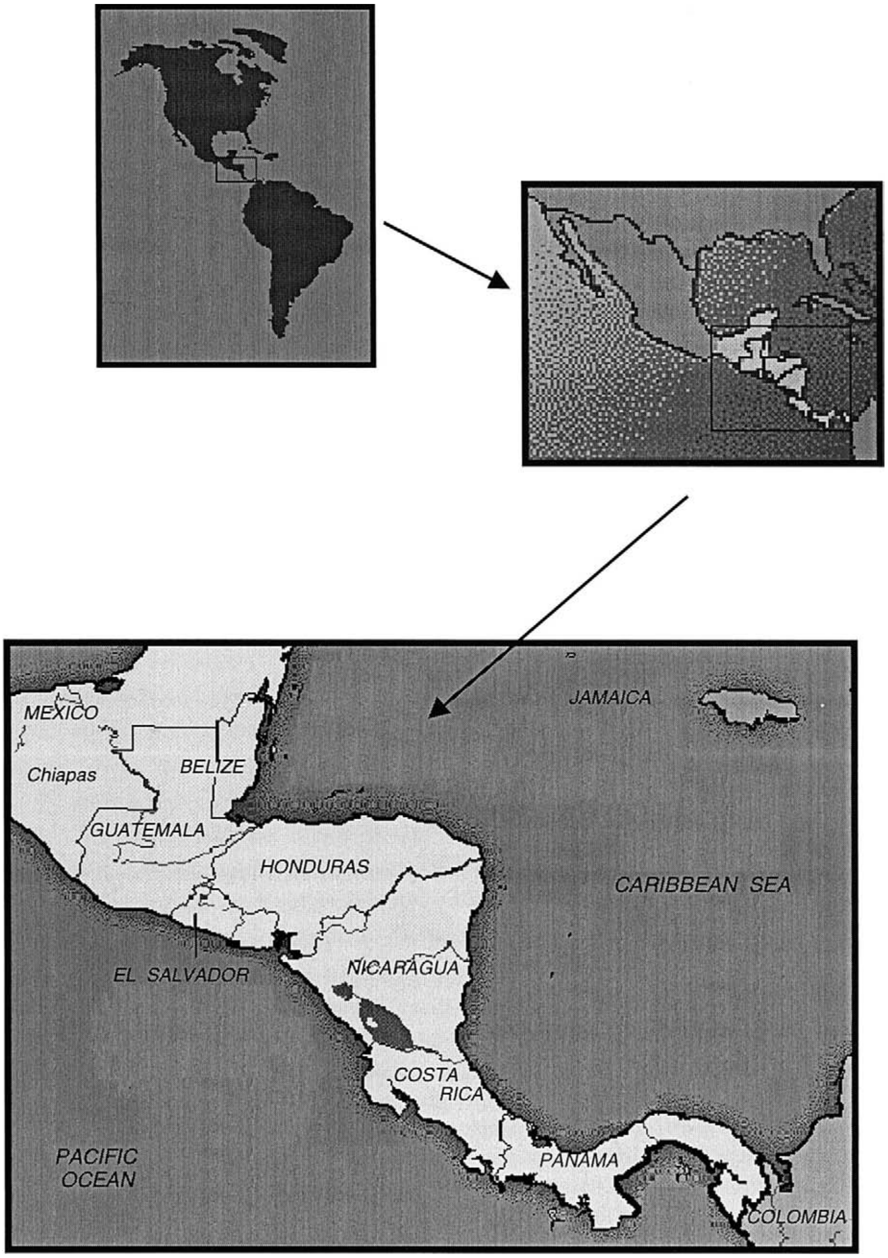


FIGURE 1 The location of Central America and the definition of the geographical region considered in this article, including the southern part of Mexico. The Mexican portion of Central America starts at the Isthmus of Tehuantepec and includes much of the state of Chiapas (see text for more details).

II. BIODIVERSITY OF CENTRAL AMERICA: AN OVERVIEW

An important facet of the biodiversity of Central America is its diversity of natural ecosystems. Because of its complex geological history and intricate topography, the region of Central America includes a mosaic of ecological conditions, in which several of the major natural ecosystems of the earth are represented. Despite the fact the region has a restricted latitudinal range, of approximately 10° (from about 8° to 18° north), and the overall climate is of a marked tropical affinity, significant climatic variations occur due to the topographic complexity rather than to latitude. Accordingly, the ecosystems of Central America are related to three major elevational/climatic situations. The lowlands, at elevations of sea level to about 900 m, with hot climates of average temperatures of 24°C, provide the conditions for the development of tropical forests, both seasonally dry and evergreen. The temperate zones, from about 900 to 1800 m above sea level (masl) and average temperatures of 18–24°C, are occupied by the exuberant tropical cloud forest and temperate zone-related coniferous and broadleaf (mostly oak) forests. The highlands, from 1800 to 3400 masl and temperatures of 12–18°C, sustain mostly high-elevation grasslands and páramos. Superimposed to this elevational gradient, the topography and the patterns of wind circulation of the region determine the existence of marked contrasts in rainfall. The cold California current of the Pacific coast chills the air, preventing it from absorbing much water vapor from the ocean and thus reducing the chances of precipitation on this side of the region. In contrast, the easterly winds of the Atlantic coast absorb much of the moisture, which is then deposited on this side of Central America as these winds flow up and over the high mountains of the region. As a result, the plains and eastern mountain slopes of the Caribbean coast receive approximately twice as much precipitation as the plains and western mountain slopes of the Pacific coast.

Such a mosaic of environmental conditions determines the existence, within a restricted geographical range, of a dramatic variety of ecosystems within a given latitudinal position. For example, within a latitudinal range of only approximately 2° and an area of approximately 51,000 km^2, the country of Costa Rica harbors 12 distinct life zones (vegetation types), one more than in the eastern United States, even though Costa Rica is only the size of West Virginia. This sample of ecosystems ranges from the seasonally dry tropical forests in the Guanacaste province to the high-elevation grasslands (“páramos”) of the central part of the country and the wet tropical forests of the Atlantic coast. At a more local scale, a southwest–northeast transect of approximately 200 km in a straight line (i.e., not considering the ruggedness of the terrain) across the mountainous region of the state of Chiapas, in southern Mexico, harbors a collection of 12 distinct vegetation types: mangroves, palm communities, savannas, seasonally dry tropical forests, tropical moist and wet forests, cloud forests, deciduous sweet gum (i.e., *Liquidambar*)–oak forests, oak forests, pine forests, grasslands, and high-elevation grasslands.

The conglomeration of natural ecosystems present in Central America is responsible for the fact that this region contains a significant proportion of the biodiversity of the planet in terms of species richness. The mere fact that the tropical rain forest, the most diverse biome of the planet, is widely distributed in the region implies that the biodiversity of Central America is of great significance. The presence of tropical rain forest is in fact one of the distinctive criteria defining the countries of megadiversity. Another important criterion for the definition of megadiversity is the presence of coastline and the ratio of coastline to land surface. The Central American coastline is approximately 5700 km (2900 km on the Pacific coastline and 2800 km on the Caribbean side); thus, there is a coastline:surface area ratio of 0.011, which is comparable to that of some of the most important megadiversity countries. Undoubtedly, marine ecosystems and the presence of coral reefs (arguably the most species-rich marine ecosystem of Earth), including some of the most important coral reefs of the Western Hemisphere (located on the Caribbean coast of Central America), add to the diversity of the region.

The well-established gradient of decrease in species richness with latitude implies that, overall, the tropical ecosystems of Central America are not expected to be as species rich as their more equatorial counterparts. The available evidence indicates that this is the case. However, the diversity of ecological conditions, and therefore of distinct ecosystems, dictates that the identity of species and species composition change considerably among localities within this restricted area. Thus, species turnover (or beta diversity) is likely to be very high in this region. The scant information available for some areas of Central America (e.g., southern Mexico) suggests that indeed this is the case. However, biological inventories are still limited and the overall species richness and its geographic variation for many plant and animal groups are poorly known for the region. Thus, it is difficult to quantitatively assess the relative contribution of beta diversity to the overall biodiversity of Central America.

In addition to the quantitative aspects of the biodiversity of Central America, a qualitative feature deserves special consideration. This refers to the combination of floristic and faunistic elements of the two major biogeographic zones that Central America is responsible for bringing together. This is more readily noticeable in the case of the flora. The elements that compose the flora of Central America belong, to a great extent, to two major floristic realms: the Arctoterciary, comprising the extra tropical territories of the Northern Hemisphere with representative genera, such as *Pinus* spp. (pines) and *Quercus* spp. (oaks), and the Neotropical, which includes the territories of tropical America, with representative genera such as *Swietenia* (caoba or mahogany) and *Cedrela* (cedro). Some Andean elements that belong to the paleooceanic realm, which comprises territories of the Andes, South America, South Africa, Australia, and New Zealand, are also represented in the Central American flora. These Andean elements include genera such as *Podocarpus* and the spectacular herbaceous plants of *Gunnera*.

The elements of the Neotropical realm are best represented in the tropical forests of the lowlands, whereas the Arctoterciary and Andean elements are predominant in the vegetation of the temperate and cold highlands. Frequently, however, there is no clear demarcation between these types of floristic elements and in several localities of Central America there is a mixture of them producing, for example, remarkable tropical rain forests at elevations of 600–900 masl, which combine the typical tall tropical trees, lianas, palms, and tree ferns with gigantic oaks and other trees of a clearly northern affinity (e.g., *Chaetoptelea*). Conversely, some cloud forests and deciduous forests, with a predominance of Nearctic elements, may include trees of clear tropical affinity such as *Cecropia* or *Nectandra*. Another remarkable feature of the Nearctic elements of Central American forests is that several of them comprise plants characteristic of the eastern United States, with a major interruption in their distribution in Texas but that penetrate into eastern Mexico and further south into Chiapas and Central America. Some of these are taxa as representative as *Liquidambar styraciflua, Nyssa sylvatica*, and *Ostrya virginiana*.

On the other hand, elements typical of the drier areas located toward the north of the Tehuantepec Isthmus are absent in the state of Chiapas (and in most of tropical Central America), but they reappear further south in the dry lands of Guatemala (Departments of Zacapa, El Progreso, and Chiquimula). This is congruent with the suggestion, presented previously, that the climatic changes leading to the recent expansion of tropical rain forests must have fragmented the distribution of elements of arid and semiarid ecosystems such as the arborescent cactus *Myrtillocactus* and trees such as *Juliania adstringens* and *Apoplanesia paniculata*.

To this amalgamation of elements of tropical, temperate, and Andean affinity, an additional contingent of endemic elements must be added. Although some elements are distributed in most of Central America, others have a more restricted distribution within the zone. For example, although several species and genera typical of the floras of Guatemala and Belize extend their distribution into Chiapas, other genera of the rain forests of these two countries are not present in Chiapas or any other tropical forests of Mexico. This also indicates that for a portion of the flora of Central America the northernmost limit of distribution is Belize and northern Guatemala. On the other hand, the most important genera of trees of the temperate and cold ecosystems of Chiapas, including *Pinus, Quercus, Liquidambar, Carpinus, Abies, Cupressus, Juniperus*, and *Taxodium*, all of which are of northern affinity, find their southernmost distribution up to Guatemala or Nicaragua. The exception is the oaks, which extend well into the south, in the Colombian Andes. Likewise, several species of oak that in all probability evolved in Central American forests (e.g., *Quercus benthami, Q. policaulis*, and *Q. crispifolia*) extend their distribution into Chiapas, which is their northernmost distribution.

A final aspect related to endemism has to do with the relatively recent discovery of new taxa as a result of biological surveys and taxonomic research in the region. In the past 10 years, two new endemic plant families have been discovered. One, the family Ticodendraceae, is a tree from Costa Rica, and the other is a herbaceous plant in the family Lacandoniaceae from the Lacandon forests in Chiapas. The latter is a truly revolutionary plant family in that the parasitic plants lack chlorophyll and the hermaphroditic flowers present the reproductive organs in a morphological disposition which is unknown among the flowering plants (male organs in the center of the flower and female organs in the periphery). Both taxa are of restricted distribution but Lacandoniaceae is known from a single population in a single locality. Further biodiversity exploration may yield additional, important taxonomic novelties.

In summary, although the degree of knowledge of biodiversity for the region is limited, the available information provides good evidence that the biological diversity of Central America is of special planetary significance. This is due not only to its quantitative aspects (e.g., diversity of species and ecosystems) but also because of the remarkable combination of life forms of different biogeographic affinity, together with endemic elements.

III. LIFE ZONES AND ECOSYSTEMS

The most synthetic and revealing way of describing the ecosystems of Central America is through the description of the vegetation. A premise for this is that the vegetation constitutes the most obvious descriptor of the ecosystem, constitutes the base of the food chain, and provides the structural matrix on which most communities and populations of animals live or indirectly depend on.

A very popular classification system is widely used to describe the ecosystems of Central America. The system, based on the life zone concept elaborated by Holdridge (1947), combines three climatic characteristics of a given region—temperature, precipitation, and altitude—and an index of the availability of moisture, potential evapotranspiration, as a predictive model to forecast a corresponding vegetation type, which is termed a life zone. Although the Holdridge system is widely used, particularly in Costa Rica, the classification system developed by Miranda and Hernández-X. (1965) is the one most used for the Mexican vegetation. This system is based on the physiognomy (i.e., the appearance: life forms and height) and phenological character (e.g., whether the predominant plants are deciduous or evergreen) of the vegetation. The conceptual basis for this system is that vegetation physiognomy reflects the adaptive response of plants to the environmental characteristics where vegetation develops. Given the facts that these two systems do not coincide in vegetation nomenclature and that the number of ecosystem types resulting from both is very large and detailed, I present a system that attempts to capture the essential aspects of the two systems and provides a more simplified classification. In addition, this simplified system attempts to present a nomenclature that is comparable to that used more generally in the scientific community. The reader is advised to consult the original systems for a more detailed description of the ecosystems of Central America.

IV. TROPICAL RAIN FORESTS

The most distinctive aspect of this type of forest in Central America is, as in other regions, the great biological diversity of species. There are many examples. The Lacandon forests of Chiapas include approximately 4300 plant species, and the La Selva field station in Costa Rica harbors 1500 plant species and approximately 411 species of birds, 116 species of mammals, and 479 species of butterflies.

In terms of their physiognomy and structure, these forests are characterized by the presence of very tall trees (30–50 m) and by their evergreen vegetation—most of the plants retain their leaves year-round. Most of the trunks of the large trees, besides being straight, have diameters at breast height of between 30 and 150 cm. However, a few, such as the *Ceiba* trees, can have even thicker trunks and their height can reach beyond the forest canopy. These so-called emergent trees provide the typical appearance of a tropical rain forest as seen from afar. Another distinctive feature of these forests is the presence of irregular or undulating contours of the tree trunks of many species, particularly at the base where they become extended protrusions, approximately triangular in profile, which may play the role of support for the tree. The structures, called buttresses, can adopt several shapes or designs that can often serve as distinctive characteristics of particular species, such as in *Dialium guianense*. The bark on the trunks, in general, is either smooth or scaly, with color ranging from light tones of *Terminalia amazonia* and some figs to dark ones of the mahogany tree and shiny reddish hues of the gumbo-limbo (*Bursera*).

Even when not clearly defined, from the canopy down there is a succession of layers of vegetation (stratification), from the subcanopy to the understory, with a spectacular profusion of life forms. The understory includes a layer of plants, most notably short palms, from less than 1 m up to 8 m. Several species of *Chamaedorea* exemplify this situation. In Central American rain forests the understory palms are typically accompanied by several species in the family Rubiaceae. At the ground level, the undergrowth is composed of ferns, several herbaceous plants such as gingers, some trailing plants, and the saplings or seedlings of the species found at higher levels.

A direct consequence of this stratification and profusion of plants is that light diminishes as one moves from the top down to ground level, where only 1–5% of the available light reaches through. Such limited availability of light leads to a series of spectacular morphological and functional adaptations in the plants to compensate for the limitation of this critical resource. Such adaptations are in turn responsible for the existence of other physiognomic and structural features of the forest, including the occurrence of a great variety of climbing plants, both woody (lianas) and herbaceous (vines, mainly in the family Araceae), together with a variety of epiphytic (i.e., plants that live on top of other plants) and semi-epiphytic herbs, shrubs, and even trees. By far, lianas are the most conspicuous climbers since they can reach over 50 cm in girth and, depending on the locality, they can constitute 1–3 of each 10 trunks. In Central America, even some species of palms have evolved the habit of liana, as is the case for *Desmoncus* and *Chamaedorea elatior*. The Araceae provide the most notable example of non-woody climbers, not only because they cover the trunks of the trees on which they climb and because of their abundance on the canopy but also because of their abundance as trailing plants at ground level, where they spread in search of trunks to climb. The vascular epiphytes, the most evident of which are the orchids, constitute a significant component of the biodiversity, given that the orchid family is commonly the most species-rich family in Central American forests. In addition to the orchids, the bromeliads constitute another important element of the epiphytic life form of these forests. Although bromeliads are not as specious, they are very abundant so that these two families constitute one of the most distinctive features of the physiognomy of the Central American rain forest.

A much less conventional type of climbing plant is that of the hemi-epiphytic trees, represented chiefly by the strangler figs, which also constitute a distinctive feature of the forests. Such plants germinate and take root on the branches of other large trees and then grow downwards at an impressive rate, producing many strong trunks that hold on the ground while covering the tree on which they established, forming a strangling network that eventually kills the initial support tree.

In addition to the impressive morphological adaptations described previously, most woody plants of the tropical rain forest have a series of ecological and functional attributes that allow them to deal with the problem of light limitation. Such attributes comprise a set of adaptations that fit the dynamic nature of the tropical rain forest. In essence, the strong winds that occur in these regions particularly during the rainiest season and winter months (November–February), overturn one or more of the giant trees of the canopy, creating so-called light gaps of up to 800 m^2. Depending on the locality, the incidence of such gap formation can be very intense, ranging from turnover rates (i.e., the number of years elapsed between two successive gaps) of 60 (e.g., in some Mexican forests) to 100 (e.g., in some Costa Rican Forests) years. The result is the production of natural clearings with high light availability and higher temperature than in the shady understory. Many plant species produce dormant seeds equipped with photosensitive mechanisms that allow them to detect the amount and type of light of the gap and germinate in these conditions. This type of plant constitutes the group of the pioneer species, the most representative of which are *Cecropia* spp. and *Ochroma*, the balsa tree. The pioneer species display a very rapid growth and recolonize the gap by growing together with some of the young plants of the typical shade-tolerant species of the mature forest that might have been growing in the understory. The gap gradually begins to close, which demands the colonization of other species physiologically adapted to increasing levels of shading. These conditions make it impossible for the survival of the pioneer species that did not manage to grow and reproduce on time. The shade-tolerant species, provided with morphological, physiological, and growth mechanisms suited to scarce illumination, make their advance to the mature forest. This process of opening and closure of the forest creates a dynamic mosaic of vegetation with varying degrees of regeneration. In the tropical rain forests of Central America the mature phase is predominant and the most representative of the described physiognomy. In addition, in this phase, the leaves of the plants are predominantly dark green and sometimes shiny on the upper surfaces, tough in texture, and with sizes that range from medium (5–10 cm in length) to large (up to 1 m or more). As in other tropical rain forests, leaves have thin, tapering extensions on their tips, which are thought to serve to drain off the excess of water. Colorful, fleshy fruits are very abundant in the plants of the mature forest, but flowers tend to be mostly small and there is a predominance of light shades of white and green; bright and shiny colors are less abundant.

The complex and diverse vegetation serves as a matrix in which a highly diverse fauna is found, particularly for insects. Given the great diversity of plants and animals, a distinctive aspect of these forests is the complex network of interactions among different species, particularly between plants and animals: pollination, dispersal (the transportation of fruits and seeds by animals), and herbivory (the consumption of plant tissue by animals). As a result, the shapes, colors, and scents of flowers and fruits and seeds, and a wealth of toxic or attractive substances, determine the behavior, feeding, and sensory patterns of animals. These animals defoliate, seek out seeds, consume nectar and other fluids, guard and defend plants against herbivores, and even attempt to copulate with orchids that resemble and smell like female bees.

V. SEASONALLY DRY TROPICAL FORESTS

This kind of forest is found in warm climates, characterized by a long dry period (from 4 to 6 months), and at altitudes from sea level up to 1500 m. As a result, in physiognomic terms they are typified by the deciduous nature of their vegetation: green and luxuriant in the rainy season and grayish and leafless when it is dry. The highest trees rarely exceed 25 m, and the dominant flora comprises smaller trees, shrubs, and long-branching trees, although vines are also found. Commonly, two vegetation strata can be defined: the canopy, with trees 15–25 m tall, and the understory, with treelets 5–15 m tall and smaller shrubs. The ground layer is very sparse in general. Although lianas can be present, the scarcity of vines and epiphytes is notable, although they can be found in microhabitats with favorable conditions and along riverbeds. Some epiphytes, such as bromeliads, can be found in abundance on trees growing on some suitable slopes. Another typical feature of these forests is the presence of numerous lichens attached to the branches of trees and shrubs, which acquire the colors of these lichens. In general, most woody plants have relatively small diameters, except in areas with greater soil moisture, such as in the vicinity of permanent or semipermanent bodies of water. As a consequence, Central American dry forests present two characteristic physiognomies: the seasonally deciduous forest associated with the hills, which is the most widespread type, and the riparian forest, smaller in extent, with larger trees that do not shed their leaves during the dry season. Frequently, riparian sectors of the dry forests include species of trees that are typically present in evergreen forests, including *Brosimum alicastrum* and several large figs. In addition to these two major variants of the dry forest, in rocky outcrops or as they extend into even drier climates or toward their northernmost distribution in Mexico, their average height decreases and there is an increased presence of thorny plants, mainly of the legume and cactus families. Another major distinctive feature of the trees in these forests is their bark, which in many species is smooth and shiny and frequently exfoliating (i.e., it peels off), as is the case of *Jatropha* or *Bursera*. In addition, several species present prominent thorns, as is the case of several members of the Bombacaceae family (*Ceiba* and *Bombacopsis*). Another distinctive aspect of the Central American dry forests is the abundance of species with small leaves, frequently of less than 5 cm in length, and many of the species present divided leaves composed of leaflets, although plants with entire or undivided leaves are also common. Several members of the legume family are representative of species with divided leaves. In addition, in marked contrast to the rain forest, dry forest plants typically present showy flowers, and many of the species produce their flowers in remarkable synchrony, particularly during the dry season when leaves are absent. These reproductive and vegetative rhythms are the norm among the plants of these forests, but notable exceptions occur, the most spectacular of which are those related to the patterns of leafing out. Although some species keep their leaves during the dry season, others such as *Jacquinia pungens* do not merely maintain their foliage during the harsh season but also lose it during the rainy season. The reasons for this leafing pattern are not fully understood by ecologists. Likewise, several species do not flower in mass or otherwise during the season of drought but rather at other times, a remarkable example of which is the *Ipomoea* tree.

The presence of endemic taxa in tropical dry forests is particularly high in these forests, in contrast to the rain forests, which share many of their species with South American forests. Endemic taxa in Central American forests include species and genera of a marked affinity with the dry lands of northern Mexico. In addition, at least one endemic life form is known to the tropical dry forests of Mexico—the *Ipomoea* tree referred to previously. This remarkable species is a member of the morning glory family, but one that has evolved as an arborescent, not a vine, life form.

Another significant contrast with the rain forest is the fact that dry forests have been studied to a much poorer extent throughout Central America. Nevertheless, the most recent review of this type of forest (Bullock *et al.*, 1995) suggests that at least some aspects of the ecological complexity of these ecosystems are comparable to those of their rainy counterparts. Many complex plant–herbivore interactions, for example, are the same as those known for rain forests. Some of them, such as the famous mutualistic interaction between the bull's thorn *Acacia* shrubs/trees and their defending *Pseudomyrmex* ants, can be even more prominent in dry forests. In this remarkable interaction, the plants provide a home (the bull's thorns) and food (sugar produced in extrafloral nectaries and specialized, lipid-rich structures) for the ants. These, in turn, are very aggressive and attack potential herbivores of the plant and even remove the vegetation in the immediate vicinity of the plant, thus avoiding potential competition.

VI. TROPICAL CLOUD FORESTS

Although with considerable variation depending on the locality, elevation, and aspect, these forests owe their physiognomy, in general, to their enormous trees. Although not very rich in species, this group of trees is very abundant. Some of these may exceed 50 m, with trunks of up to 2 m in diameter at breast height, such as is the case of some oaks and the spectacular *Talauma* *mexicana* in the cloud forests of Chiapas. In the mature forest areas the cloud forest is clearly dominated by woody plants. Among these, the lianas, although not completely absent, are considerably scarce. A few of the trees lose their leaves during the driest time of the year, but because a high proportion of the trees do not lose their foliage or, if they lose it, they replace it very quickly, the forest as a whole can be regarded as an evergreen ecosystem. Although the cloud forests located in the southern parts of Central America are dominated by species of tropical affinity, toward their northern distribution, particularly in Chiapas, elements of boreal affinity become prevalent and intermingle with the tropical elements. This is the case for the oaks (e.g., *Quercus candicans* or *Q. skinnerii*) and especially for the sweet gum tree *Liquidambar styraciflua*. This noticeable tree, of approximately 50 m, extends in the Central American cloud forests from Honduras to Chiapas. Although the dominant trees of the more southern forests, such as in Costa Rica, are mostly 30–40 m tall, and a stratification is more clearly defined than in their more northern counterparts, the rest of the physiognomic characteristics are very similar. Although buttresses are present, they are not as developed as in the tropical rain forest trees. The understory and lower strata of the cloud forest are notable for the great profusion of palms and ferns, the most conspicuous of which are the spectacular tree ferns. In some areas these can reach up to 20 m. The ferns, including smaller species, tree ferns, and their immature forms, achieve a degree of abundance in the cloud forests that is hardly comparable to that of any other Central American ecosystem. Understory trees can be 8–15 m tall and frequently have leaning, crooked trunks and relatively long crowns. The shrub layer is fairly dense and can reach 2 or 3 m. The ground layer is considerably covered by ferns, *Selaginella*, and several broad-leaved herbs. Bluish coloration in several plants of these lower strata is quite common. Another distinctive feature of these forests is the great profusion of epiphytes and semi-epiphytic plants. Most of the trees are covered by dense associations of ferns, moss, orchids, bromeliads, peperomias, herbaceous vines, and other vascular plants such as the gesneriads. A spectacular fauna of characteristic birds, including quetzals, tanagers, and horned guans, accompanies such variety of plants.

Perhaps the most extensive cloud forest of Central America is that of Costa Rica. Nevertheless, this type of forest is of a very restricted extension and patchy distribution. As a result, and because of the relatively poor knowledge of its ecology and also because of its crucial role in retaining water, this Central American ecosystem requires attention and formal protection before it is converted into plantations and grasslands.

VII. TEMPERATE FORESTS

Temperate-like forests of Central America include two major types (oak forest and pine forest), but in some parts, in which the dominant species overlap, a mixed oak–pine forest can be distinguished. Oak forests range from the tropical montane and premontane forest dominated by *Quercus* spp. and Lauraceae in Costa Rica to the clearly oak-dominated forests ("encinar") of Chiapas, Mexico. The Costa Rican oak forests are evergreen communities of low to intermediate height, with trees 20–25 m tall and unbuttressed trunks and rough bark. The understory is open, with trees 4–15 m tall and with slender trunks and highly ramified crowns, with some tree ferns present. The understory is dominated by bamboo, particularly in the disturbed areas. The branches and trunks of trees are heavily covered with herbaceous epiphytes and moss. The oak forests are best represented in Chiapas, where they form distinct associations ranging from woodlands intermingled with the tropical rain forest (with *Quercus oocarpa*) or the cloud forest (*Q. candicans* and *Q. skinneri*) to dense mono-dominated oak forests at elevations of approximately 3000 m, with trees 35 m in height (*Q. acatenanguensis*). The most widespread oak forests of Chiapas are those of intermediate elevations (700–2500 masl). In the drier areas near the central depression of the state, these are low forests with grasses in the understory and they are sometimes adjacent to seasonally dry forests and savannas. In contrast, in areas of higher elevation and where winds are still loaded with water vapor, they are similar to cloud forests: dense, and loaded with epiphytes, particularly bromeliads and orchids. The most impressive forests of this type are those near the spectacular Montebello Lagoons. Other associations include the low oak forest of 2–4 m in height, which resembles a chaparral, near the city of Comitán.

The pine forests have their greatest distribution in the territory of Chiapas and extend, more sparsely and with fewer associations and diversity, toward the south to Honduras. With the exception of some fragmentary information for the pine forests of Chiapas, these Central American ecosystems are very poorly known scientifically. In Chiapas, these forests occupy most of the surface of temperate and cold areas, to which pines are very well adapted. This is due to the reduced surface of their leaves (needles), which are also well protected on their surface, allowing them to resist long periods of drought and low temperatures. In addition, the resin present in these plants' tissue provides resistance to several pests and the thick bark of the trunk and branches help to reduce the effects of fire. In general, these forests present a clear dominance of a single tree species. They extend from intermediate elevations up to the timberline at approximately 4000 m and where rainfall does not exceed 1200 mm per year. However, even in areas of more precipitation, pine forests can be occupied by species with thin and flexible needles (*Pinus strobus* and *P. ayacahuite*). The most widespread pine forests, at elevations of 750–3000 m, are dominated by *P. oocarpa*, whereas *P. hartwegii* and *P. rudis* dominate the more restricted pine forests of the coldest regions (at elevations of 2900–4000 m).

In most cases, the understory is dominated by dense grassland. A significant, although seldom considered, aspect of the biodiversity of both oak and pine forests is the presence of a rich flora of shrubs and herbs, many of northern affinity but including several species endemic to these ecosystems.

This variety of Central American oak and pine forests, of peculiar and restricted distribution, is very poorly studied ecologically and from a biodiversity point of view, but these forests are disappearing at alarming rates, and most of what remains is heavily fragmented. It is likely that some significant aspects of their ecology and biodiversity have been very strongly disrupted or lost.

VIII. OTHER HIGH-ELEVATION ECOSYSTEMS

Two distinct ecosystems are present at the highest elevations in Central America: the páramo and the high-elevation grassland or zacatonal. The latter is an association of tall and dense grasses that extend in the clearings of pine and oak forests of the highest elevations. This indicates that their current extension is most likely secondary; that is, derived from human perturbation due to the considerable resistance of the grasses to the fires that sweep these regions. The height of these systems is 0.5–2 m and the dominant grasses include *Epicampes macrura, Festuca amplissima*, and *Stipa ichu*. The Central American páramo is the northernmost occurrence of the Andean páramo. In Costa Rica it is dominated by shrubs where drainage is adequate and by bogs where drainage is poor. Another portion of páramo occurs in Chiapas, at elevations of 4000 m or higher (above the timberline of the Tacaná Volcano), where it is physiognomically similar to the zacatonal but the dominant grasses (*Calamagrostis* and *Festuca*) are shorter, sometimes forming a prairie. Frequently, the distance among grasses is considerable and there is much open space for the establishment of other plants, many of which barely rise above the level of the ground. This group of plants includes several members of the rose family (*Alchemilla* and *Potentilla*) and the daisy (*Senecio* and *Gnaphalium*) and cabbage (*Draba*) families. The small size and prostrate habit of the plants of this ecosystem are advantageous. They are able to take advantage of the heat of the layer of air close to the ground during daylight since the higher layers, even during daylight, are extremely cold and incompatible with their functioning and survival.

IX. CENTRAL AMERICAN ECOSYSTEMS IN LIGHT OF GLOBAL ENVIRONMENTAL CHANGE

Several aspects of global environmental change are predicted to have a profound effect on the biodiversity of the planet. Among these, climatic change has received the greatest attention, but recent information indicates that, in addition, the current patterns of land use (leading to deforestation and habitat fragmentation) may significantly affect biodiversity. Although information is limited, particularly for this region, it is of interest to review what the available models and information suggest regarding the situation of Central American ecosystems in the light of climatic change and land use patterns.

A. Climatic Change

General circulation models have been developed to forecast the changes in temperature and rainfall and their geographic variation under a given set of assumptions, mostly related to CO_2 concentrations. These models suggest that the greatest changes of climate are to be expected at higher latitudes. Consequently, given the relatively low latitude of Central America, it could be expected that the natural ecosystems of this region would be affected to a minor degree. Indeed, the maps resulting from these models depict Central America as a region of little or no climatic change. Although this is an optimistic result, it is important to bear in mind that such maps and forecasts are, by necessity, of very gross resolution. Unfortunately, no detailed analyses are available to assess the situation at a finer geographical scale. Nevertheless, a recent study on the expected changes in vegetation coverage resulting from climatic change in the forests of Mexico (Villers and Trejo, 1998) provides some useful insights, given that the study included the northern part of Central America (Fig. 1). In addition, the study considered the major Central American ecosystems described previously (i.e., tropical rain forest, tropical dry forest, cloud forest, pine, and pine–oak temperate forest). Ecosystem-response scenarios were developed with two models based on a doubling of CO_2 concentration (the Geophysical Fluid Dynamics Laboratory and Canadian Climate Center models) and under the assumption of a homogeneous temperature increase of 2°C and 10% reduction in rainfall (Fig. 2). In synthesis, the results show that, in comparison to their current potential distribution, tropical forests of the warm lowlands would be the least affected: Seasonally-dry forests are expected to remain about the same, whereas rain forests are even predicted to increase. In contrast, temperate forests appear to be affected the most: Pine forests are predicted to disappear, whereas cloud forests would experience reductions of 45–75%. Moreover, the results were very consistent among the different models. In summary, whereas global models forecast little climate change impact on the ecosystems of the region, more detailed models suggest that this is likely the case for the tropical ecosystems of the warm/lowland climates but the temperate ecosystems are expected to be significantly impacted.

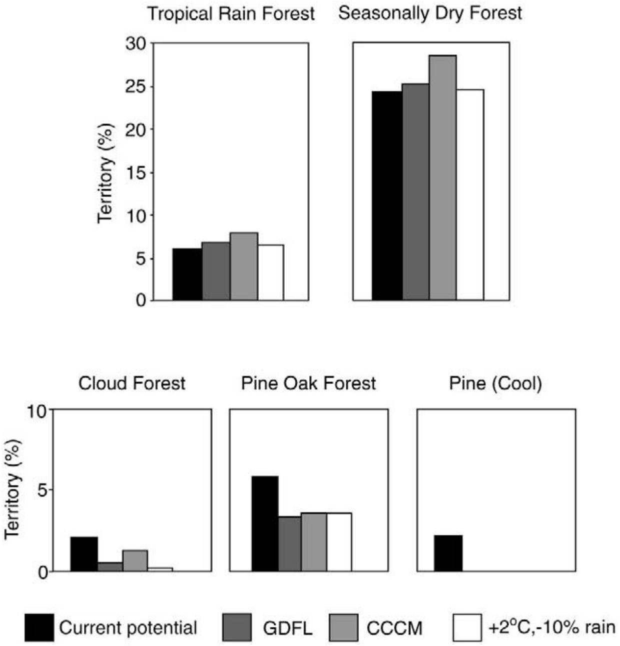


FIGURE 2 The expected changes in the coverage of five Mexican forest ecosystems under three scenarios of climate change in comparison to their potential current coverage (% of the territory of the country). The scenarios of climate change were derived from two models that assume a doubling of atmospheric CO_2 (GDFL and CCCM) and under the assumption of a homogeneous increase of 2°C and a 10% decrease of rainfall. The absence of bars in the chart corresponding to the cool pine forests indicates that this ecosystem is expected to disappear completely under the three scenarios of climate change (modified from Villers and Trejo, 1998).

B. Land Use Patterns

The best available descriptive index of the patterns of land use is the rate of deforestation. Detailed studies on the rates of deforestation have been published only for some regions of Central America, particularly Costa Rica and the Mexican portion of the region. Studies on deforestation in the tropical rain forest of Lacandonia, Chiapas (Mendoza and Dirzo, 1999) show a considerable degree of variation in deforestation rates among localities, with a range of 0–8% and an average of 1.6% per year. Even if these values seem low, the estimated absolute deforestation rate for this area is 6286 ha per year.

The situation in Costa Rica is particularly revealing, given that it is a country that has made a significant effort to conserve its natural ecosystems. Despite this effort, the available statistics suggest that deforestation has been extremely high in this country. For example, studies from the early 1990s indicate that most of the country's ecosystems are heavily deforested and the percentages of area of habitat lost range from 32 (montane rain forest) to more than 90% (dry deciduous forest, lowland moist forest, premontane moist forest, and lower montane moist forest). The aggregated value for 11 of the 12 ecosystems (life zones) of the country is reported to be approximately 80%.

These isolated studies give the impression that deforestation in Central America is considerable and an aggregated analysis (World Resources Institute, 1991) confirms such impression and permits comparison of the overall situation of the region with that of other parts of world (Fig. 3). The estimated amount of percentage of forest converted to nonforest use in Central America is about 3.5 times greater than the corresponding estimate for the whole world and more than twice the value for the forested ecosystems of South America and Asia.

In synthesis, the information presented in this section indicates that even if climatic change will not

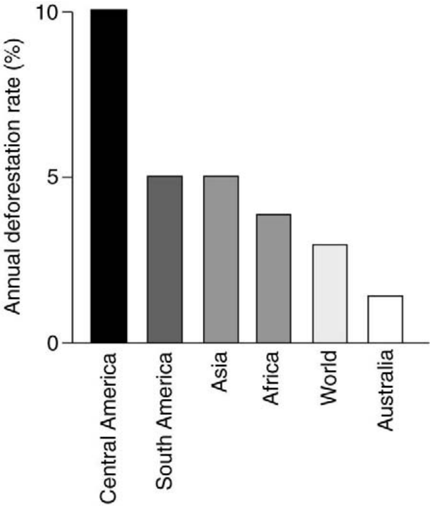


FIGURE 3 The estimated annual rate of deforestation (% forest area lost per year), as an indicator of the annual change in forest coverage, for five regions, including Central America and the world as a whole (modified from World Resources Institute, 1991).

have a major impact on the tropical forests of Central America, forests of all types seem to be seriously threatened by the current patterns of land use. Moreover, the available information suggests that such threats might compensate for the potentially positive effects, or exacerbate the negative effects, of climatic change. Finally, the current patterns of land use and the expected effects of climatic change on the temperate forests might have a significant effect on the functioning of the ecosystems and overall structure and composition of the biodiversity of Central America. The relevant information to assess the consequences of the loss or degradation of the Central American ecosystems is very limited. In addition, such an analysis is beyond the scope of this article. Nevertheless, our knowledge of their biodiversity and the threats they are currently experiencing indicates that a significant effort toward their conservation and wise management is a major agenda for the scientific community and society at large.

See Also the Following Articles

BIODIVERSITY-RICH COUNTRIES • NORTH AMERICA, PATTERNS OF BIODIVERSITY IN • RAINFOREST ECOSYSTEMS, ANIMAL DIVERSITY • RAINFOREST ECOSYSTEMS, PLANT DIVERSITY • SOUTH AMERICA, ECOSYSTEMS OF

Bibliography

Bullock, S. H., Mooney, H. A., and Medina, E. (Eds.). (1995). *Seasonally Dry Tropical Forests*. Cambridge University Press, Cambridge.

Holdridge, L. R. (1967). *Life Zone Ecology*. Tropical Science Center, San Jose, Costa Rica.

Janzen, D. H. (Ed.) (1983). *Costa Rican Natural History*. Univ. of Chicago Press, Chicago.

McDade, L. A., Bawa, K. S., Hespenheide, H. A., and Hartshorn, G. S. (1994). *La Selva: Ecology and Natural History of a Neotropical Rain Forest*. Univ. of Chicago Press, Chicago.

Mendoza, E., and Dirzo, R. (1999). Deforestation in Lacandonia (southeast Mexico): Evidence for the declaration of the northernmost tropical hot-spot. *Biodiv. Conserv.*, **8**, 1621–1641.

Miranda, F. (1998). *La Vegetación de Chiapas* (3rd ed.), Consejo Estatal para la Cultura y las Artes de Chiapas. Tuxtla Gutiérrez, Chiapas.

Miranda, F., and Hernández-X., E. (1956). Los Tipos de Vegetación de México y su Clasificación. *Bol. Soc. Bot. Méx.* **28**, 29–178.

Rich, P. V., and Rich, T. H. (1983). The Central American dispersal route: biotic history and paleogeography. In (D. H. Janzen, Ed.). *Costa Rican Natural History*, pp. 12–34. Univ. of Chicago Press, Chicago.

Rzedowski, J. (1978). *Vegetación de México*. Limusa, Mexico City.

Villers, L., and Trejo, I. (1998). Assessment of the vulnerability of forest ecosystems to climate change in Mexico. *Climate Res.* **9**, 87–93.

World Resources Institute (1991). *World Resources Report 1991–1992: A Guide to the Global Environment*. Oxford Univ. Press. New York.

CLADISTICS

Ian J. Kitching, Peter L. Forey, and David M. Williams
The Natural History Museum, London

GLOSSARY

apomorphy A derived character or character state; if two or more taxa share apomorphies, these are referred to as synapomorphies.

clade Group of taxa diagnosed as monophyletic by the discovery of homologies (or synapomorphies).

cladogram Branching diagram specifying hierarchical relationships among taxa.

cladogram support Tests that permit some evaluation of how well data fit a cladogram.

consensus cladogram (tree) Branching diagram that summarizes the common branching patterns from two or more cladograms.

homology Two characters passing the similarity, conjunction, and congruence tests are termed homologous; in cladistics, homology is synonymous with synapomorphy.

homoplasy A character or character state acquired by parallel or convergent evolution that bears resemblance to a character in a different group.

monophyly Relationship between taxa united by a synapomorphy.

optimization Procedure for reconstructing the most parsimonious sequence of character change on a cladogram.

parsimony General scientific principle that given alternative explanations or hypotheses for a set of observations or data, the most corroborated is that requiring the fewest ad hoc (ancillary or additional) hypotheses.

plesiomorphy An apomorphic character or character state that specifies a more inclusive group than that under consideration.

CLADISTICS is a class of methods of biological classification that groups taxa hierarchically into discrete sets and subsets. This article presents the principles and concepts of cladistics and describes the principal analytical methods. The operations by which observations of organisms are coded for analysis are explained, followed by the methods for reconstructing the hierarchical relationships among taxa (usually expressed as branching diagrams termed cladograms). Statistics and principles for determining the degree of fit between data and cladograms are discussed, which permit choices to be made among competing cladograms.

I. RELATIONSHIPS

The basic concept of cladistics is that genealogical connections among organisms are expressed in relative terms. Consider three taxa, A, B, C, whose genealogical relationships are as given in Fig. 1a. Taxa B and C are more closely related to each other than either is to taxon A because they share a common ancestor, x (which lived at time t_1), that is not shared with taxon A. Similarly, taxon A is more closely related to the group (B + C) because A, B, and C together share a unique common ancestor, y, that lived at an earlier time (t_0). In a real example (Fig. 2), the human and turkey are considered to share a unique common ancestor (w) that lived at t_3. Similarly, the frog, turkey, and human are more closely related to each other than to either the perch or dogfish because these three taxa uniquely share a common ancestor x that lived at time t_2. The human and turkey are called sister-groups. Likewise, in this example, the frog is the sister-group of (human + turkey). The aim of cladistic analysis is to infer the sister-group hierarchy of life-forms by analysis of characters and to express the results as branching diagrams. These diagrams are called "cladograms" because they identify a hierarchical arrangement of taxa based on homologies termed "clades."

A. Types of Characters

Two types of characters are recognized based on where they occur in the inferred phylogenetic history of a group (Fig. 1b). The character that occurs in the ancestor is termed "plesiomorphic" (near to the ancestral morphology) and the derived character is "apomorphic" (away from the ancestral morphology). Apomorphic and plesiomorphic are relative terms, that is, relative to a particular systematic problem. In Fig. 1b, character a′ is apomorphic with respect to character a, but plesiomorphic with respect to character a″.

Cladistic analysis proceeds by identifying shared apomorphic characters or "synapomorphies." In Fig. 2, a four-chambered heart and endothermy are synapomorphies that suggest the human and the turkey share a unique common ancestor w. The cladogram implies that these two characters arose in ancestor w and were then inherited by both the human and the turkey. Synapomorphies may therefore be considered as evolutionary homologies. In contrast, the shared possession of internal nostrils and pentadactyl limbs by the human and turkey does not imply that they share a unique common ancestor because these attributes are also found in the frog. These shared primitive characters (or "symplesiomorphies") are inherited from an ancestor more remote than the most recent common ancestor of the human and the turkey. They are thus irrelevant to the hypothesis of a relationship between the human and the turkey. However, with respect to the more inclusive three-taxon problem comprising the frog, turkey, and human, internal nostrils and pentadactyl limbs are relevant. At this level, they are synapomorphies suggesting that these three taxa form a group with a common ancestry at x. Apomorphies occurring in only a single terminal taxon are termed "autapomorphies." In Fig. 2, these are prismatic cartilage (dogfish), spiny

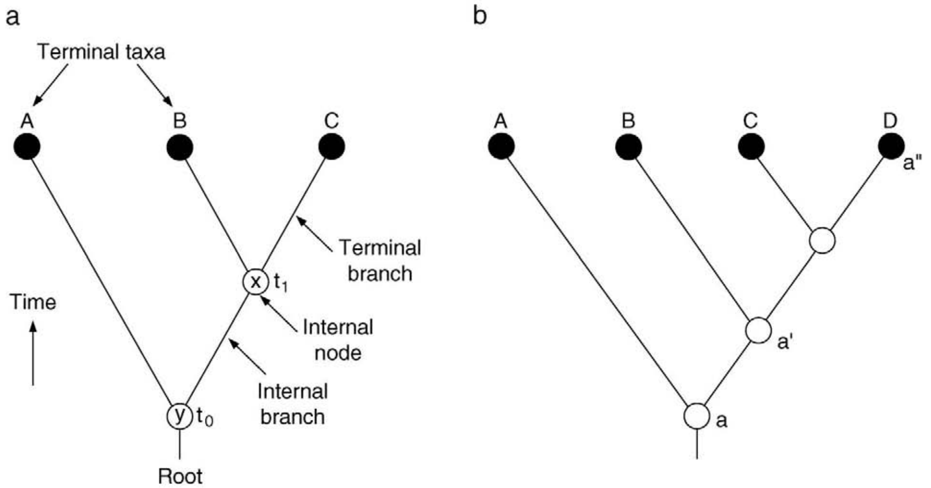


FIGURE 1 (a) The phylogenetic tree and the cladistic meaning of relationship. Taxa B and C are considered more closely related to each other than either is to taxon A because they share a unique common ancestor (x) that is not shared by taxon A. (b) Plesiomorphy and apomorphy are relative terms. On this phylogenetic tree, a′ is apomorphic with respect to a but plesiomorphic with respect to a″.

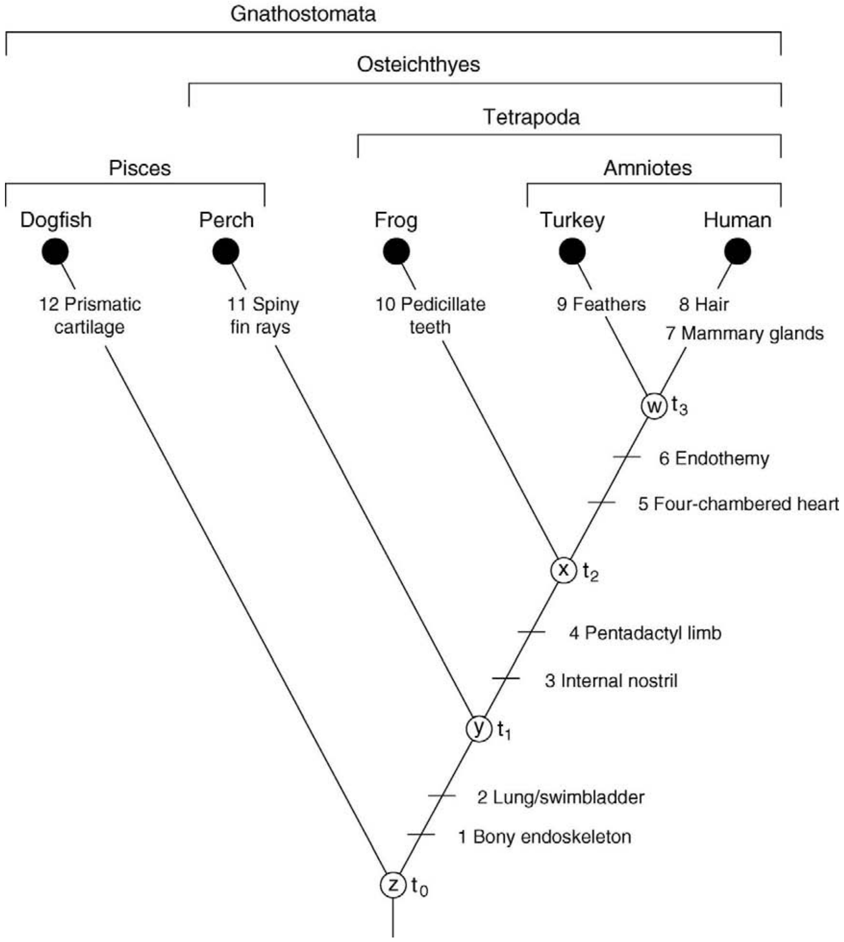


FIGURE 2 A phylogenetic tree for five taxa of vertebrates. Three monophyletic groups are established using characters 1–6, while autapomorphic characters 7–12 diagnose the terminal taxa. The group Pisces is paraphyletic because one of its included members (the perch) is cladistically more closely related to Tetrapoda.

fin rays (perch), pedicillate teeth (frog), feathers (turkey), and hair and mammary glands (human). However, if a terminal taxon is itself a group, then its autapomorphies are also synapomorphies of its component taxa.

Thus, as with the cladistic meaning of relationship, characters are also relative, depending on the systematic problem under consideration. Furthermore, it should be stressed that characters are observations of the features occurring in organisms and the explication of their hierarchical distribution need not imply a particular theory of evolution.

B. Parsimony

Relationships among three taxa (as in Fig. 1a) can be resolved in three ways—A (B C), B (A C), and C (A B)—whereas for four taxa (as in Fig. 1b) there are fifteen possible fully resolved cladograms. In cladistic analysis, parsimony is the universal criterion for selecting among alternative hypotheses of character distribution. Characters are fitted onto alternative topologies and the cladogram that accounts for the greatest number of characters in the simplest way is chosen as the best hypothesis of relationships.

Suppose six characters are distributed among four taxa as shown in the taxon/character matrix in Fig. 3. Taxon A has none of the characters but the other three taxa each have a different complement. Characters 2 and 4 are autapomorphies because they are each present in only one of the taxa. They are uninformative for grouping taxa (they serve only to diagnose the terminal taxa). Characters 1, 3, 5, and 6 are potentially useful because they occur in more than one taxon. Given the

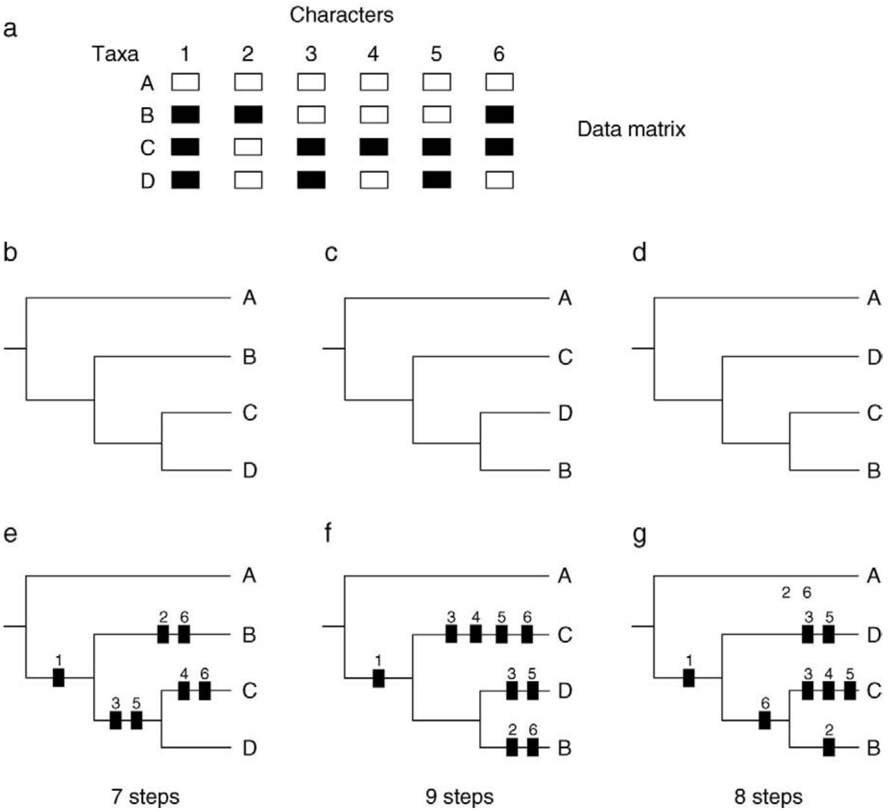


FIGURE 3 Parsimony. (a) A data matrix of six characters distributed among four taxa, A–D. Plesiomorphic states are shown by open boxes and apomorphic states by black boxes. Taxon A is totally plesiomorphic. (b–d) The three possible resolutions of taxa B–D relative to A. Placing the characters on these three topologies results in (e) being selected as the optimal cladogram while (f) and (g) are suboptimal.

three taxa that have potentially informative information, there are three ways in which these taxa can be arranged dichotomously (Fig. 3b–d). If the characters are now placed onto each possible cladogram, according to the groups they specify, then three different results are obtained (Fig. 3e–g). In Fig. 3e, all characters except one appear only once. In this solution, character 6 must be assumed to appear twice, once in taxon B and once in taxon C, which are not sister-groups. In this example, the behavior of character 6 is homoplastic, that is, it occurs more than once on the cladogram and is said to be a homoplasy. In contrast, in the other two topologies (Fig. 3f and 3g), we must assume that two or more characters appear more than once. Hence, the cladogram in Fig. 3e accounts for the distribution of the characters in the most economical way and is thus the preferred solution.

The distribution of characters can also be regarded as the number of steps on a cladogram, which, in Fig. 3, is the number of instances where a character is gained. In Fig. 3e, this is seven, while the other cladograms (Fig. 3f and 3g) are more costly, requiring nine and eight steps, respectively. The concept of steps is actually a little more subtle than the sum of character gains because a character may appear at one point on a cladogram and then disappear at another point. For example, another explanation for the distribution of character 6 in Fig. 3e is to assume that it is gained by the group (B + C + D) and then lost in taxon D. Each change, whether gain or loss, is considered a step. In this example, both accounts of character change demand two steps and therefore both hypotheses of character change are equally parsimonious. The sum of the number of steps on a cladogram is termed the length of the cladogram, irrespective of whether the changes are gains or losses. The most parsimonious solution is also known as the optimal cladogram and the other cladograms (i.e., those requiring more than the minimum number of steps to explain the character distributions) as suboptimal.

It is possible for a given set of characters to yield two or more equally most parsimonious cladograms. Then, we may prefer to accept one of the solutions based on other criteria, such as a closer agreement with the stratigraphic record or by differential weighting of one type of character change relative to another. Alternatively, we may simply accept that the conflict in the data is such that we cannot derive a unique most parsimonious solution. For certain purposes, we may choose to combine those components common to the different solutions to form a consensus cladogram.

C. Groups

Cladistics recognizes only monophyletic groups of organisms, which are those based on synapomorphies. Monophyletic groups are the only groups that can be circumscribed by objective boundaries. In evolutionary terms, monophyletic groups comprise the most recent common ancestor and all of its descendants. In Fig. 2, Amniota, Tetrapoda, Osteichthyes, and Gnathostomata are all monophyletic. Two other types of "groups" are sometimes referred to but these are not groups in the same sense as monophyletic groups. Paraphyletic "groups" are based on symplesiomorphy; in evolutionary terms, their members are linked by common ancestry but one or more of the descendants of the most recent common ancestor are excluded. In Fig. 2, Pisces (fishes) is a paraphyletic assemblage. Many taxa traditionally regarded as ancestral, such as fishes, reptiles, and green algae, are paraphyletic. Polyphyletic "groups" are based on homoplasy, that is, characters that are considered convergently derived and that cannot be inferred to have been present in the most recent common ancestor of the included taxa. In Fig. 2, an assemblage comprising the dogfish and the turkey (perhaps based on the observation that both lay eggs surrounded by a shell, although no one would claim such a homology) would be a polyphyletic group.

D. Cladograms and Phylogenetic Trees

A cladogram is a diagram that summarizes a pattern of character distribution. Usually, a cladogram is drawn as a branching diagram (e.g., Fig. 1). The nodes denote a hierarchy of synapomorphies but there is no necessary implication of ancestry and descent. Cladograms may also be written in parenthetical notation or illustrated as a Venn diagram (Fig. 4a), which conveys the same grouping information as a branching diagram. In contrast, phylogenetic trees include a time axis and embody concepts of ancestry and descent with modification. In phylogenetic trees, the nodes denote ancestors (known or hypothetical) and the branches imply character change. Several phylogenetic trees may be compatible with the pattern of character distribution implied by a cladogram (Fig. 4b). Some of these trees allow the possibility that one or more taxa are ancestral to others. Only the phylogenetic tree that assumes all nodes represent hypothetical ancestors has the same topology as

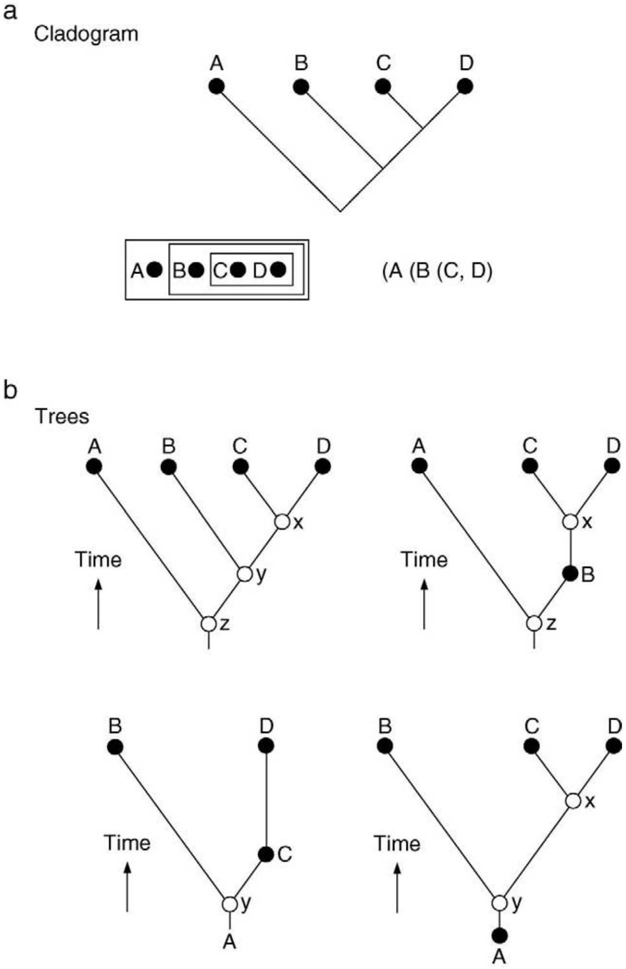


FIGURE 4 Cladograms and trees. (a) A cladogram simply shows the hierarchical distribution of characters. It has no time axis and may be drawn as a branching diagram, a Venn diagram, or in parenthetical notation. Given this cladogram, several phylogenetic trees may be inferred from the same data. Some of these are shown in (b). All except that on the top left invoke the concept of one of the taxa being ancestral. The tree at the top left assumes only hypothetical ancestors and is the only one that is synonymous with the cladogram.

the cladogram. Thus, cladograms are more general than phylogenetic trees, which are precise statements about ancestry and descent.

II. CHARACTERS AND CODING

Opinions differ over the nature and discovery of taxonomic characters. One view holds that characters are properties of organisms that provide quantifiable variation. Alternatively, characters may be viewed as theories concerning two (or more) attributes, which may look different but are nevertheless considered the same. This latter view is embraced within a general understanding of homology, such that characters may be understood in the same manner as homology. There is a lack of agreement over what indicates the discovery of a character. However, all definitions of homology suggest that it concerns features that are similar in different taxa sharing a recent common ancestor. Such definitions satisfy as explanations but do not aid discovery.

A. Homology

Within cladistics, various tests assist in establishing homology. One view of characters is that they are identical in meaning and discovery to homology, and homology may be conceived as a series of three tests that apply to methods of comparison.

The similarity "test" suggests that without evidence to allow direct comparison of one feature with another, there would be no proposition of homology and, consequently, no concept of a character. This "test" is not exact and cannot be taken to imply "identity." For example, comparisons may consider the detailed similarity of any two stamens or the inferred similarity of mammalian stapes with gnathostome hyoid arches.

The conjunction test suggests that two features that co-occur in the same organism cannot be considered homologous. A familiar example, albeit contrived, is an angel with both wings and forearms. The two kinds of limbs in the same individual cannot be considered homologous. Many comparisons fail this test and are often associated with "homomony" or serial homology (e.g., the individual vertebrae of a single vertebral column or the abdominal appendages of arthropods).

The congruence test is considered the most exacting and refers to the support afforded to one homology by others. In other words, homologies are considered to have passed the test if there are other homologies that specify the same taxon. Congruence is actually an analytical procedure and is usually considered in terms of parsimony. However, it also points to another property of homology, that is, homology can never be proven. As data are accumulated, previously supported homologies may be overturned and new theories of homology established in their place.

B. Character Recognition

It is generally agreed that characters, however conceived, are based on observations. Stated simply, a feature (e.g., stamens, shoulder girdles, wings) is observed in a particular specimen and directly translated into the character. This approach may initially seem useful and would eventually lead to enumeration of all features of the specimen. However, the final list would not consist of "characters" but would be an inventory of "features," each being a descriptive element of the specimen and implying that such descriptions apply to all specimens of the same taxon. Each descriptive element contains a notion of theory.

Suppose the specimen examined is a rat. Initially, it would be straightforward to describe: head, body, limbs, tail, and so on. More detailed examination would reveal, for example, a vertebral column. We identify the vertebral column by drawing on knowledge of previous studies of rat anatomy and are able to confirm its detailed similarity to other vertebral columns. In so doing, we assimilate what is already known of vertebrates: they are animals with a vertebral column. If this process were performed for all features, then it would seem that all attributes of this single rat could uncover its place in the hierarchy of life, at every inclusive level. In this sense, taxonomic characters are very much like homologies: features that (potentially) specify a particular taxon. A vertebral column does not tell us it is a rat; it tells us that it is a vertebrate. In this sense, the vertebral column is a *feature* of any particular rat but only a *character* of vertebrates. This distinction identifies the general task of systematics: to identify the level at which various attributes are homologies (characters) diagnosing taxa.

C. Kinds of Characters

Characters are often thought of as comprising different types. Some refer to different numbers of a feature, and others refer to differences in structure. For example, variation in stamen structure in angiosperms encompasses both different forms of anthers and filaments and differences in their numbers. This exemplifies the distinction between quantitative and qualitative characters, the former usually being counts ("meristic characters") or measurements ("biometric characters"), the latter relating to structural differences. Quantitative characters may be problematic for cladistics insofar as it can be difficult to render measurements and counts as meaningful homology statements. This is not to say that such characters are not useful, for they can serve to identify particular specimens. However, their use may be limited because they are not always amenable to cladistic analysis.

D. Characters as Phylogenetic Evidence

Although structural evidence is sought for cladistic purposes, the observed features themselves are not necessarily the characters. For example, some organisms have fins, others have arms, and yet others have wings. Studies of fins, arms, and wings show that they have certain parts in common as well as certain parts that are unique. These common properties might suggest an initial proposition that fins, arms, and wings are all examples of a single character, in this case "paired appendages." However, further details are needed to confirm this hypothesis.

Suppose the "wings" examined were from an insect such as a housefly. It might still seem reasonable to call them "paired appendages," but in this example there is nothing in the detailed anatomy to suggest that housefly wings and mammalian arms (or even bird wings) are in any way "the same." One conclusion is that the term "wings" is ambiguous when describing attributes of organisms but not when describing the function of these attributes (here, flight). Wings may indeed be considered as a part of an organism but flight is their assigned function (usually). A more reasonable conclusion is that "wings" is not a character at all but a functional attribute. Thus, the wing of a bird is better considered as a modified "paired appendage," modified for flight. The problem, however, is not simply semantic. If the comparison was made between a bird and a bat, then detailed anatomy does suggest that both are indeed "paired appendages" and that both are modified for flight. Our current understanding of vertebrates suggests that birds and bats do not form a monophyletic group. Hence, bird wings and bats wings would be considered two characters rather than one when interpreted on the cladogram (Fig. 5). Their "sameness" is captured as "forelimbs," their differences as wings of birds and wings of bats.

Consideration of the wings of birds, the forelimbs of tetrapods, and the fins of fishes together identifies a well-defined character ("paired appendages") with various manifestations. These manifestations might suggest particular taxonomic groups. There might be a taxon with "fins," a taxon with "arms," and a taxon with "wings." This, of course, was precisely the situation for many years: fishes (Pisces) have fins, tetrapods have arms, and birds (Aves) have wings (Table I). Derived from these observations is the implication that fins, wings, and arms are in some way connected, other than by being "paired appendages":

fins–arms–wings

Furthermore, that connection might be viewed in an evolutionary context, such that it represents the transformation of one manifestation (e.g., fins) into another (e.g., arms):

fins → arms → wings

In more general terms, features could be represented as modifications of other features, such that they are hierarchically related:

Fins
Standard fins (fins)
  Modified fins (arms)
    Modified arms (wings)

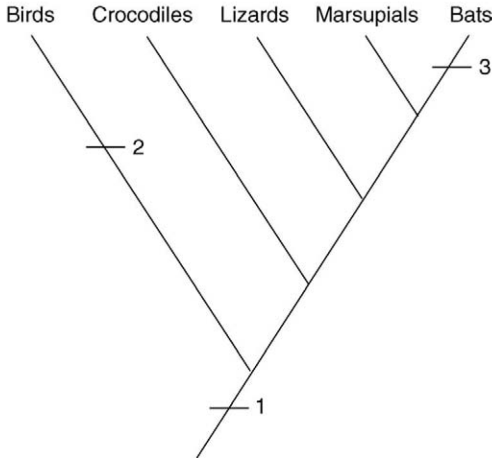


FIGURE 5 Relationships of vertebrates; 1 ≡ "forelimbs"; 2 ≡ "wings of birds" 3 ≡ "wings of bats."

TABLE I

Taxon	Character
Pisces	Fins
Tetrapods	Arms
Birds	Wings

The features are no longer structured as a series of alternatives, as in the first example (fins–arms–wings), but now specify a nested set of relationships. The proposition is that wings are really kinds of arms, and that arms are really kinds of fins, and fins represent the entire set of animals with "paired appendages." It is possible to interpret all of these taxa (fishes, tetrapods, and birds) as having fins. Hence "fins" is not a character of fishes but rather of gnathostomes (in this example, fishes + mammals + birds). Consequently, both the character "fins" and the taxon Pisces disappear. This view confirms the notion that characters are hypotheses drawn from observations rather than simply the observations themselves. Such hypotheses are identical to those made for general statements of homology. Superficially, the relationship between fins, arms, and wings may be considered identical to "fins → arms → wings," as was implied by Hennig (1966) in his concept of "transformation series." However, it is possible to view characters as more general, specifying particular relationships in terms of a definitive statement connecting to a taxon.

E. Character Coding

One significant outcome of theories relating to characters is how they might be represented numerically for cladistic analysis. For example, one might code each "character" (fins, wings, arms) separately (Table II, characters 2–4). This scheme reflects the "uniqueness" of each attribute but contains no information relevant to recognizing that the three observed forms are connected as "paired appendages." This approach is referred to as "absence/presence binary coding," because a positive value (usually 1) is assigned to the presence of a feature and a negative value (usually 0) is assigned to the absence of the feature. Alternatively, one might represent the same series of observations in a single column to signify their connection (as "paired appendages"), then assign each unique feature a separate value (Table II, character 1). This is "multistate coding" and considers the character to be composed of discrete states that bear some (usually unspecified) relationship to one another. Hence different values appear in the same column and are treated as dependent on the other values. This might not be seen as completely sufficient, as additional information would be needed to specify the exact nature of the connection. For instance, one might wish to specify that the "characters" are connected but that the nature of that connection is unknown. Choices of this nature relate to character optimization (see Section III,C).

TABLE II

	Characters			
Taxon	1	2	3	4
Fins	1	1	0	0
Wings	2	0	1	0
Arms	3	0	0	1

To summarize, characters have their origin, but not their identity, in observations. Characters are what lead us to suspect that taxa exist (vertebral column ≡ vertebrates ≡ taxon Vertebrata) and hence are identical to conjectures of homology derived from empirical investigation of specimens. Homology is the relation that specifies taxa and that implies an intimate relationship between characters and taxa. Both are the results of analyses and are discovered by our investigation of features.

III. CLADISTIC ANALYSIS

A. Cladogram Construction

The original method of cladogram construction was proposed by Hennig (1950, 1966) and is thus known as Hennigian argumentation. In this approach, characters are first polarized into plesiomorphic and apomorphic states. The groups thus diagnosed by synapomorphies are then organized manually into a cladogram. However, this procedure can only find the most parsimonious cladograms when the data are free or nearly free of homoplasy (i.e., the fit of data to most parsimonious cladogram is perfect or nearly so). For larger and more complex data sets, computerized algorithmic methods become a necessity.

There are two main computerized approaches to cladogram construction. Exact methods guarantee to find the most parsimonious cladograms. The simplest exact method is "exhaustive search." First, three taxa are chosen and connected to form the only possible unrooted, fully resolved cladogram for these taxa. Then, a fourth taxon is selected and added to each branch to yield the three possible fully resolved, partial, unrooted cladograms for four taxa. A fifth taxon is then selected and added to each of the five branches on these three partial cladograms to yield the fifteen possible fully resolved unrooted topologies for five taxa. This process is continued, following every possible path of taxon addition, until all taxa have been added and all possible fully resolved cladograms have been found. The lengths of these cladograms are then calculated and the shortest is chosen as the optimal solution(s). However, as the number of taxa increases, the number of cladograms to be examined rises exponentially and the time required for exhaustive search soon becomes unreasonable.

One exact method that does not require every possible cladogram to be evaluated is "branch-and-bound" analysis. In this approach, a preliminary cladogram is constructed and its length is set as the upper bound for subsequent searches. A procedure similar to an exhaustive search is then undertaken but at each step the length of the partial cladogram is recorded. Whenever this length exceeds the current upper bound, that partial cladogram is rejected (and so, consequently, are those complete topologies that would be derived from it by adding the remaining taxa). By this means, the number of topologies to be examined is reduced. Once all taxa have been added, the length of the complete cladogram is examined and if it is equal to the upper bound, then that topology is retained as a most parsimonious cladogram. However, should this cladogram be shorter than the current upper bound, then its length is substituted as a new upper bound. This important procedure allows subsequent partial cladograms to be rejected quickly and thus speed up analysis. This process continues until all possible paths have been examined, whence the set of optimal cladograms will have been found.

For large data sets (more than 30 taxa), even branch-and-bound analysis can be too time-consuming. In this case, approximate or "heuristic" methods are used. These approaches examine only a subset of all possible topologies and thus are not guaranteed to find the most parsimonious cladogram(s). However, they are faster than exact methods for large numbers of taxa and thus certainty of finding the optimal cladogram(s) is sacrificed for decreased computational time.

Heuristic analysis comprises two stages. In the initial building phase, a cladogram is constructed using a process of "stepwise addition." The order in which taxa are added is termed the "addition sequence" and there are various ways in which taxa may be added. Once a complete cladogram has been constructed, attempts can be made to improve upon it by performing a series of rearrangements called "branch-swapping." The cladogram is cut into two or more partial cladograms, which are then recombined in order to try to find new, shorter topologies. The efficiency of current branch-swapping algorithms in finding most parsimonious cladograms is very high, but it is always possible that they can become trapped in a local optimum. Thus, one should always be aware with heuristic analyses that shorter topologies than those reported may exist.

B. Character Polarization and Cladogram Rooting

Manually implemented cladistic methods, such as Hennigian argumentation, require synapomorphies to be identified in advance of cladogram construction. The process through which plesiomorphic and apomorphic characters are distinguished is termed "character polarization." Numerous criteria for polarizing characters have been proposed, but only two are now considered valid.

The first criterion, called "outgroup comparison," was classified by Nelson (1973) as an "indirect" method, because it draws upon evidence from a source (the "outgroup") that is external to the taxa under investigation (the "ingroup"). In its most basic form, polarization using outgroup comparison can be defined as follows: "For a given character with two or more states within a group, the state occurring in related groups is assumed to be the plesiomorphic state" (Watrous and Wheeler, 1981: 5). This definition is adequate when all outgroup taxa share the same state, but it is insufficient if the outgroup taxa are heterogeneous. Maddison *et al.* (1984) further noted that it is inappropriate to estimate the state in the most recent common ancestor of the ingroup (the "ingroup node"). Rather, it is the state at the next most distal node, linking the ingroup to the first outgroup (the "outgroup node"), that should be estimated if the solution is to be globally optimal, and they described an algorithmic approach to such reconstruction.

In contrast, Nelson (1973) classified the "ontogenetic criterion" as a "direct" method, because its implementation relies on evidence from the ingroup taxa alone. It is defined as follows: "Given an ontogenetic character transformation from a character observed to be more general to a character observed to be less general, the more general character is primitive [plesiomorphic] and the less general character derived [apomorphic]" (Nelson, 1978: 327). For example, the embryos of both sharks and frogs have cartilaginous skeletons. However, this condition persists into the adult shark, but in frogs the cartilage is replaced by bone during ontogeny. In this example, a bony skeleton is observed to be less general (occurring only in the frog) and is thus interpreted as apomorphic. In this context, "more general" is defined as that occurring earlier in ontogeny.

As such, the more general character is not simply the more common (although this may often be the case) and the ontogenetic criterion does not equate with commonality. What is important is that the less general character is nested within the observed distribution of the more general character. This requirement is violated by ontogenetic sequences that are secondarily simplified through paedomorphosis or neoteny. Such an ontogeny cannot be interpreted as proceeding from the more general to the less general, and Nelson's criterion will not allow us to distinguish a secondarily reduced ontogeny from the plesiomorphic sequence. Then, we rely on congruence with other characters to make the distinction.

Most recent cladistic studies do not actually include a priori polarization of characters, but undertake "simultaneous, unconstrained analysis": "simultaneous" because both ingroup and outgroup taxa are analyzed together, and "unconstrained" because outgroup taxon relationships are unspecified before analysis. Cladograms are then rooted between the outgroup node and the remaining outgroup taxa, at which point character polarities are established.

C. Optimization

Optimization is the process of determining the sequence of character state changes on a cladogram in order to test hypotheses of transformation. If the data include characters coded as multistate, then these may be interpreted according to many different optimality criteria, of which the best known are Wagner and Fitch optimization.

Wagner optimization (Fig. 6a–b) is used for "ordered" or "additive" multistate characters, in which

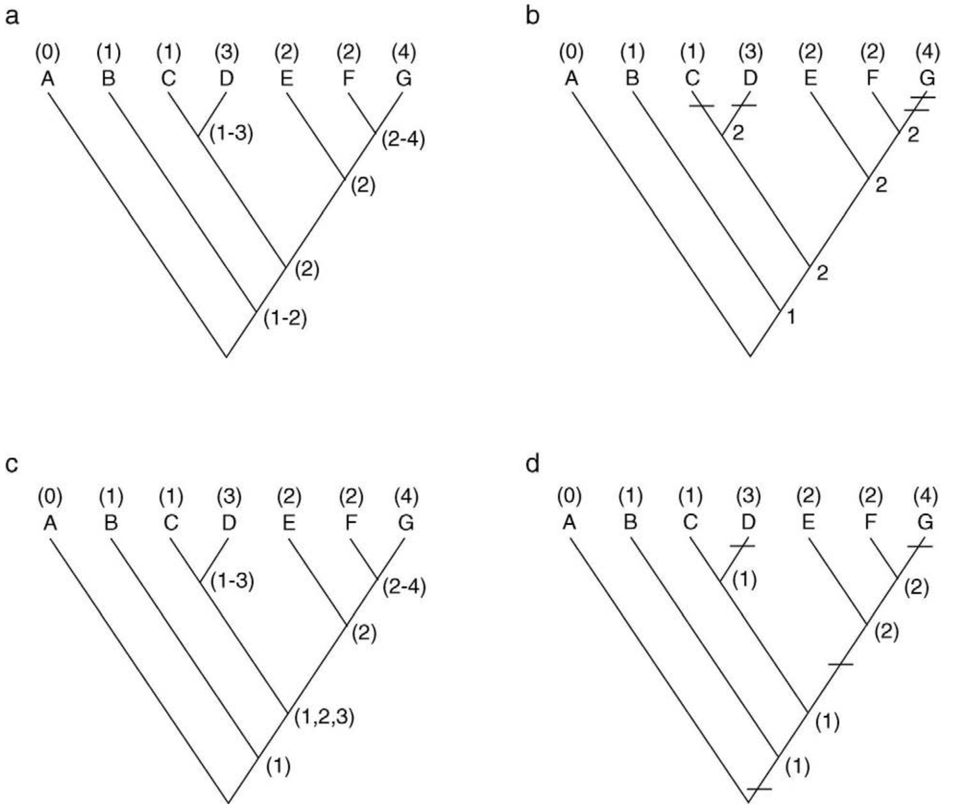


FIGURE 6 Character optimization. (a) Wagner optimization (ordered multistate characters). State sets are assigned to internal nodes on the first pass from the terminal taxa toward the root. (b) States are assigned to the internal nodes on the second pass from the root to the terminal taxa. (c) Fitch optimization (unordered multistate characters). State sets are assigned to internal nodes on the first pass from the terminal taxa toward the root. (d) States are assigned to the internal nodes on the second pass from the root to the terminal taxa.

transformations between successive states are considered as incremental. Thus, the changes 0 ↔ 1 and 1 ↔ 2 each "cost" the same number of steps (usually one), whereas the change 0 ↔ 2 is considered to "pass through" state 1 and thus costs two steps. Costs are symmetrical, so that the changes 0 → 1 and 1 → 0 both constitute a single step (termed "free reversibility").

Wagner optimization is implemented in two stages. First, the minimum number of steps for a character is determined. On a pass down the cladogram, from the most distal taxa toward the root, each of the internal nodes is assigned a "state set" (Fig. 6a), which is defined as the intersection of the two derivative state sets. If the intersection is empty, then the smallest closed set that includes an element of each derivative state set is assigned. For example, the intersection of the state sets of taxon F (2) and taxon G (4) is empty. Thus, the smallest closed set, (2–4), is assigned to the node linking these two taxa. In contrast, the intersection between this state set and that of taxon E is not empty. Both contain state 2, and this value is assigned to the node joining taxa E, F, and G. Unambiguous states are then assigned to internal nodes by a second pass, up the cladogram (Fig. 6b), to produce a "most parsimonious reconstruction." If the state set of a node contains more than one value, then the node is assigned the value that is closest to that of the node of which it is a derivative. Thus, the node joining taxa F and G is assigned state 2, because this is the value of the next most distal node. The most parsimonious reconstruction has six steps because the character is ordered. As a result, the change from state 2 to state 4 along the branch leading to taxon G counts as two steps.

Fitch optimization (Fig. 6c–d) is used for "unordered" or "nonadditive" multistate characters in which transformations between any two states are considered equal. Thus, the changes 0 ↔ 1, 1 ↔ 2, and 0 ↔ 2 all cost a single step. As with Wagner optimization, costs under Fitch optimization are freely reversible.

Fitch optimization follows similar procedures to Wagner optimization but with two differences. First, the state set assigned to an internal node is the union of the two derivative state sets (Fig. 6c). Second, when determining the most parsimonious reconstruction, a node with an ambiguous state set is assigned the value of the next most distal node if that value is an element of the ambiguous state set. Otherwise, an element is selected arbitrarily. This most parsimonious reconstruction (Fig. 6d) has four steps because the character is unordered. As a result, the change from state 1 to 3 along the branch leading to taxon D and the change from 2 to 4 along the branch leading to taxon G each count as a single step.

These procedures do not necessarily yield a unique most parsimonious reconstruction. For example, it is equally parsimonious to optimize the ordered character using the nodal state reconstructions shown in Fig. 6d for the unordered character. State 1 is assigned to the node joining taxa C–G, and is followed by a two-step change, 1 → 3, on the branch leading to taxon D and a one-step change, 1 → 2, on the branch joining taxa E–G. The steps on the branches joining taxa B–G and leading to taxon C would be lost, thus maintaining the most parsimonious length of six steps. This type of optimization, in which changes are placed onto the cladogram as far from the root as possible, is called "delayed" or "slow" transformation. In contrast, "accelerated" or "fast" transformation places changes onto the cladogram as close to the root as possible, as in Fig. 6b. When alternative most parsimonious reconstructions are possible, character optimization may lead to "spurious resolution," in which groups appear to be resolved but have no unambiguous support in the data. Such groups are not strong hypotheses of relationship, and Nixon and Carpenter (1996) suggested that they should be eliminated wherever possible (without violating the minimum length requirement). Those cladograms that remain, which are both of minimum length and have all groups supported unambiguously by data, are termed "strictly supported cladograms" and are the preferred topologies.

IV. CLADOGRAM EVALUATION

A. Character Fit

The preferred solution to a cladistic analysis is the most parsimonious cladogram because this represents the simplest explanation of the data with the number of ad hoc hypotheses of homoplasy kept to a minimum (see Fig. 3e). Consequently, the most basic measure for assessing the fit of data to a cladogram is cladogram length. The most parsimonious cladogram has best fit because it is the shortest; longer cladograms have poorer fit. However, the length of the most parsimonious cladogram is partly dependent on the absolute size of the data set from which it is derived. Larger data sets will necessarily yield longer cladograms than smaller data sets. A binary character will display perfect fit when it is placed on a cladogram with a single step. Homoplasy is manifest as an increase in the number of steps. The amount of homoplasy implied by a character on a cladogram is measured by the "consistency index" (ci), which is the ratio of the minimum number of steps required by the character ($m = 1$ for a binary character) to the observed number (s). In Fig. 3e, character 5 occurs only once and hence its ci = 1 ($m/s = 1/1$), whereas character 6 shows two steps and thus its ci = 0.5 ($m/s = 1/2$). The amount of homoplasy implied by the whole data set can be measured using the "ensemble consistency index" (CI), which is the ratio of the minimum number of steps implied by all characters (M) to the length of a cladogram (S). For the cladogram in Fig. 3e, the CI = 0.86 ($M/S = 6/7$).

There are three perceived problems with the consistency index as a measure of homoplasy. First, although uninformative characters do not add any grouping information to a cladogram, they will inflate its CI. However, this is of significance only when different data sets are being compared. Second, CI can never attain a zero value. A data set in which all possible informative characters occur in equal numbers (an "undecisive" matrix) provides no evidence for preferring one cladogram to any other. Nevertheless, these cladograms will all have positive, nonzero CI values. Third, it has been observed empirically that CI decreases as the number of taxa increases, irrespective of change in the information content of the data. However, this is a recognized and expected property of the CI.

Although the consistency index is useful as a measure of the amount of homoplasy in a character or data set, it is indifferent to the pattern of fit on a cladogram. A binary character that occurs on two separate terminal branches will have the same ci value (0.5) as one that supports two separate groups of taxa. However, in the former case, the character contains no grouping information, while in the latter it is a synapomorphy (albeit homoplastic) for two groups of taxa. The amount of similarity in a character that is interpreted as synapomorphy is measured by the "retention index" (ri). This is defined as $(g - s)/(g - m)$, where s and m are the same variables as for ci, and g is the maximum number of steps that a character can show on any cladogram. For character 5 in Fig. 3e, the minimum and observed number of steps is one, and the maximum number is two. Hence its ri = 1 $[(g - s)/(g - m) = (2 - 1)/(2 - 1)]$ and all similarity is interpreted as synapomorphy. In contrast, for character 6, the minimum number of steps is one, and the observed and maximum number is two. Hence its ri = 0 $[(g - s)/(g - m) = (2 - 2)/(2 - 1)]$ and none of the similarity is interpreted as synapomorphy. The method can be extended to the whole data set as the "ensemble retention index" (RI), which uses the summed values of g, s, and m (G, S, and M, respectively). For the cladogram in Fig. 3e, the RI = 0.67 $[(G - S)/(G - M) = (9 - 7)/(9 - 6)]$.

B. Character Weighting

The application of differential weights to characters has a long history in systematics. Methods of weighting can be divided into a priori and a posteriori procedures, depending on whether they are applied before or after cladogram construction.

A priori approaches to character weighting generally invoke beliefs that some characters are more important than others or use a particular model of evolution or character change, under which certain types of transformation are considered more or less likely than others. For example, it is common when analyzing nucleotide sequence data to downweight transition substitutions relative to transversions. Alternatively, changes in third codon positions may be disregarded (i.e., accorded zero weight) because they are considered much more likely than changes in first or second positions as a result of the redundancy of the genetic code. Numerous other models have been proposed and they are particularly frequent in the field of molecular systematics. However, such weighting schemes are justifiable only insofar as their underlying model is justifiable.

A posteriori weighting schemes are based on "cladistic consistency" (i.e., the fit of characters to a cladogram) and characters with greater fit are accorded greater weight. One indication of a character's fit is the amount of homoplasy it shows on a cladogram. However, homoplasy does not imply that all similarity is uninformative and the proportion of similarity interpreted as synapomorphy also needs to be taken into account. Hence, both the consistency index (homoplasy) and the retention index (synapomorphy) can be used to estimate character weights. Farris (1989) suggested using the product of these two measures, the "rescaled consistency index" (rc). By combining the ci and ri in this way, characters in which none of the similarity is synapomorphic (ri = 0) receive zero weight, irrespective of their level of homoplasy. All other characters, which contain some amount of grouping information (ri > 0), are differentially weighted according to their level of homoplasy. Using this approach, in Fig. 3e character 5 would receive a weight of 1 (ci = ri = 1), but character 6 would receive a weight of 0 (ci = 0.5, ri = 0). These weights are applied in a new analysis and the most parsimonious cladogram(s) obtained are used to estimate a new set of weights. This procedure is repeated until a stable set of both weights and most parsimonious cladograms is achieved; hence the name "successive approximations character weighting."

The level of homoplasy of a character may also be viewed as the number of extra steps required to fit it to a cladogram. If all extra steps in all characters are considered equal, then a linear fitting function is being applied to their relative cladistic consistency. For example, in Fig. 3e, the single step of character 5 is considered equal to each and either of the two steps of character 6. However, intuitively, we might consider that characters showing fewer extra steps are "better" than those showing more. The former can be assigned higher weights than the latter using a concave fitting function of relative cladistic consistency. This approach was implemented by Goloboff (1993) as "implied weighting," in which the weight (W) accorded to a character is calculated as $W = K/(K + ESi)$. ESi is the number of extra steps shown by the character and K is the "constant of concavity." The value of K can be varied to weight more or less strongly against those characters with the most extra steps. As K decreases, these characters will receive progressively lower weights. For example, in Fig. 3e, character 6 will receive a weight of 0.85 when $K = 6$, but a weight of only 0.5 when $K = 1$ (the "perfect" character 5, which has no extra steps, receives the maximum weight of 1). The optimal cladogram is that for which the summed weights for all characters has the largest value.

C. Consensus Cladograms (Trees)

A cladistic analysis will often produce more than one most parsimonious cladogram as a result of contradictory signal in the data (homoplasy). The agreement between such "fundamental cladograms" (so-called because they are generated directly from the analysis of data) can be conveniently summarized by means of a consensus cladogram (usually referred to as a consensus tree). Several consensus methods have been proposed, of which the most widely used are "strict," "combinable components" (or "semistrict"), "Adams," and "majority-rule."

The strict consensus tree is the most conservative, because it includes only those groups (often referred to as "components") that are common to all the fundamental cladograms. For example, in the two cladograms shown in Fig. 7a and 7b, only groups ABC and DEF occur in both, and thus these are the only groups that appear in the strict consensus tree (Fig. 7d). Groups EF, AB, and BC are excluded because the first is lacking from Fig. 7b (where it is unresolved) and the other two are contradictory.

However, it is possible for a group to be lacking from one or more fundamental cladograms and yet be uncontradicted. For example, group EF in Fig. 7a does not conflict with the cladogram in Fig. 7b because it is one of the three resolutions possible for the group DEF. Combinable components consensus allows such non-replicated, but non-conflicting, groups to be included in the consensus tree, in addition to those groups in common (Fig. 7e). When all fundamental cladograms are fully resolved, with no spurious resolution due to ambiguous optimization, then the strict and combinable components consensus trees will be the same.

A problem with both strict and combinable components consensus is that a single taxon appearing in highly disparate positions on two cladograms is sufficient to collapse all intervening resolution. For example, taxon D in Figs. 7a and 7c appears as the sister-group to two different terminal taxon-pairs. Consequently, the strict consensus tree (Fig. 7f) is relatively unresolved. However, examination of Figs. 7a and 7c shows that taxon D is acting as a "rogue" taxon; that is, apart from its differing positions, the resolution of the remaining taxa is identical in the two cladograms. Such rogue taxa can be identified using Adams consensus analysis. On an Adams consensus tree, taxa in conflicting positions on the fundamental cladograms are placed at the most inclusive node they have in common; in other words, the consensus contains all intersecting sets of taxa common to the fundamental cladograms. However, as a result, it is possible for an Adams consensus tree to contain groups that are not found on any of the fundamental cladograms and thus they need to be interpreted with care. Sometimes taxa such as D would simply be deleted to give the "largest common pruned tree" as a consensus.

When the number of fundamental cladograms is large, strict consensus trees are often very poorly resolved and can be viewed as too restrictive. One method of increasing resolution of the consensus is to retain those groups that occur in a prespecified number of the fundamental cladograms. Typically, such majority-rule consensus trees will comprise those groups that occur in more than 50% of cladograms. The majority-rule consensus tree of the cladograms in Fig. 7a–c, shown in Fig. 7h, is fully resolved, despite the marked topological differences in its fundamental cladograms.

Regardless of their number, the most parsimonious cladograms found by cladistic analysis remain our best estimate of the relationships among the taxa under study. Because resolution is lost, most consensus trees

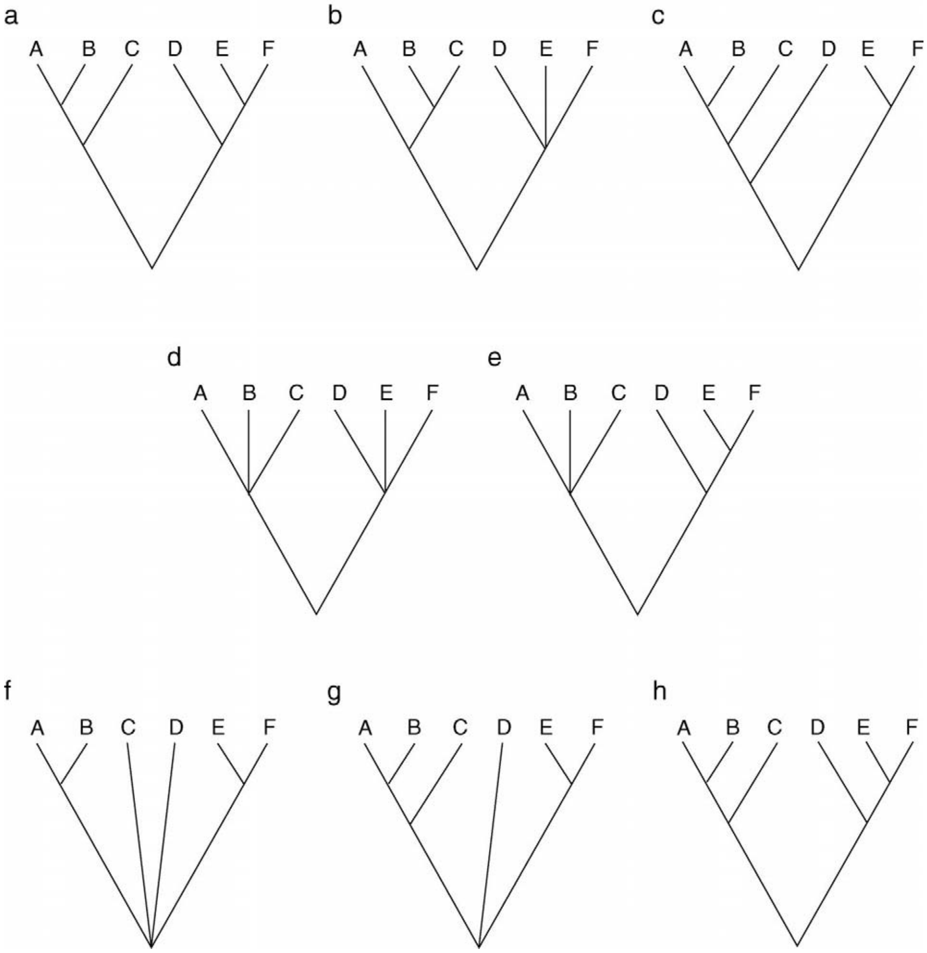


FIGURE 7 Consensus analysis. (a–c) Three cladograms for six taxa, A–F. (d) Strict consensus tree of cladograms 7a and 7b. (e) Combinable components consensus tree of cladograms 7a and 7b. (f) Strict consensus tree of cladograms 7a and 7c. (g) Adams consensus tree of cladograms 7a and 7c. (h) Majority-rule consensus tree of cladograms 7a, 7b, and 7c.

are less parsimonious than their fundamental cladograms. However, if the topological differences among the fundamental cladograms are due solely to ambiguous optimization, the strict consensus tree will also be of minimum length. Then, the strict consensus tree is also the strictly supported cladogram (Nixon and Carpenter, 1996) and is the preferred most parsimonious topology because it is the only cladogram that is both of minimum length and has all groups supported unambiguously by data.

D. Cladogram and Group Support

A number of statistics attempt to assign levels of support or confidence to the results of cladistic analyses. They can be divided into two categories: methods that address support for an entire cladogram and aim to determine whether there is any "significant" structure in the data, and methods that examine the support afforded to individual groups on a cladogram and attempt to distinguish those groups that are well supported from those that are not.

Methods aimed at assessing support for an entire cladogram all use the same general principle. The length of the most parsimonious cladogram obtained from the observed data set is compared with those derived from a large number of "phylogenetically uninformative" data sets, with the expectation that the former will be substantially shorter than any of the latter. Several definitions of "phylogenetically uninformative" data have been proposed, including “statistically random” (scores in a data matrix are allocated at random), “undecisive” (all possible informative characters occur in equal numbers), and “randomly co-varying.” The last of these forms the basis of the “permutation tail probability” (PTP) test. Covariation here is the degree to which characters are explicable by the same cladogram (i.e., congruence). A most parsimonious cladogram that contains a large amount of homoplasy may be derived from data exhibiting such poor covariation that randomly co-varying characters could produce a cladogram of equal length or shorter. The PTP test uses pseudoreplicate data sets in which character codes are randomly reassigned to taxa, with the restriction that the proportion of each code is maintained, and with each character treated independently. For example, for three taxa (A, B, and C) originally coded 0 and two (D and E) coded 1, permutation may reassign 0 to A, C, and D and 1 to B and E. This procedure is repeated to create a large number of such pseudoreplicates, for which the most parsimonious cladograms are then found. The PTP is defined as the proportion of all data sets (original plus permuted) that yield most parsimonious cladograms at least as short as the original data set and may be interpreted as the probability that a cladogram of this length could have arisen by chance.

The simplest measure of support for a particular group on a cladogram is branch length. However, homoplasy makes the assessment of branch support difficult and groups may appear to be better supported than they actually are. “Bremer support” is a more precise measure of clade support and is defined as the number of extra steps required to lose a clade from the strict consensus tree of near-minimum-length cladograms. When there is no homoplasy in the data, the Bremer support of a group is the same as its branch length. Otherwise, support is reduced to the extent that there are alternative equally parsimonious groupings. To calculate Bremer support, first those cladograms that are one step longer than minimum are found and the strict consensus tree formed from them and the most parsimonious cladogram. The process is repeated, adding a step at a time, until the group in question is lost from the consensus. The number of extra steps required to achieve this is the Bremer support for the group. If more than one most parsimonious cladogram is found initially, then the procedure begins with the consensus tree of these cladograms. Any group that may be a potential resolution of the consensus will have a Bremer support equal to 0.

The bootstrap seeks to estimate group support using pseudoreplicate data sets, which are formed by randomly sampling characters with replacement. The effect is to weight some characters and delete others, with the constraint that the total weight equals the original number of characters. A large number of such pseudoreplicates are generated and their most parsimonious cladograms are found. Conflict among these cladograms is assessed using a majority-rule consensus tree and the support for a group is estimated as the proportion of most parsimonious cladograms on which it is recovered.

However, both the PTP test and bootstrap are questionable in a cladistic context. The PTP test’s null hypothesis of randomly co-varying characters is contrary to the basis of cladistics. Characters as nested homology statements are intrinsically hierarchical and thus it is inappropriate to measure their performance against randomized characters from which this hierarchy has been removed. The bootstrap assumes that a data set represents a random sample of all possible characters. However, taxonomic characters are generally carefully selected with the aim of resolving the relationships of the taxa under study. Furthermore, it only requires a single synapomorphy to diagnose a clade. However, the random nature of the bootstrap process means that such a character may be represented in only a few pseudoreplicates and thus the group will not appear in the majority-rule consensus tree despite being uncontradicted. Thus, bootstrap values are a one-sided test; recovered groups have some measure of support in the data, but groups that are not recovered cannot be rejected.

V. SIMULTANEOUS AND PARTITIONED ANALYSIS

It is generally recognized that data from many sources may be used in a cladistic analysis (e.g., morphological, physiological, behavioral, ecological, or molecular sequences) and that analyses of these data can yield different hypotheses of relationships. Simultaneous analysis (sometimes called a total evidence approach) combines all data, from whatever source, into a single taxon × character matrix for analysis. The resulting hypothesis of relationships is thus determined by character congruence (Fig. 8a). Alternatively, different classes of data may be analyzed separately and the resulting cladograms added together using a consensus method to extract the common phylogenetic signal. This is called the partitioned evidence approach (Fig. 8b) and the result is determined by taxic congruence. The reasoning

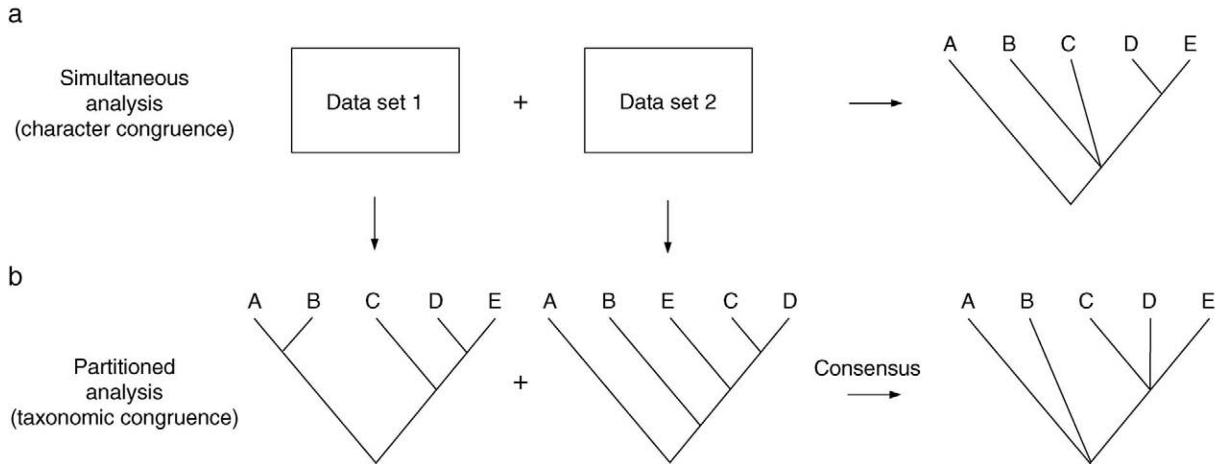


FIGURE 8 Simultaneous and partitioned analysis. In simultaneous analysis, all data are combined into a single matrix before analysis. In partitioned analysis, each data set is analyzed separately and the resulting cladograms are then “added” together using a consensus method.

underlying this approach is that different classes of data may reflect different evolutionary processes and so should be analyzed separately. A third alternative, known as conditional data combination, attempts to distinguish those conditions under which it would be best to keep data sets separate and conduct partitioned evidence analysis from those conditions under which it is more appropriate to conduct a simultaneous analysis. This approach estimates the degree of heterogeneity of phylogenetic signal among data sets, and if the heterogeneity is greater than might be explained by sampling error, then the data sets are analyzed separately.

See Also the Following Articles

CLADOGENESIS • DIVERSITY, TAXONOMIC VERSUS FUNCTIONAL • EVOLUTION, THEORY OF • SYSTEMATICS, OVERVIEW • TAXONOMY, METHODS OF

Bibliography

Farris, J. S. (1989). The retention index and the rescaled consistency index. *Cladistics* **5**, 417–419.

Goloboff, P. A. (1993). Estimating character weights during tree search. *Cladistics* **9**, 83–91.

Hennig, W. (1950). *Grundzüge einer Theorie der phylogenetischen Systematik.* Deutsche Zentralverlag, Berlin.

Hennig, W. (1966). *Phylogenetic Systematics.* University of Illinois Press, Urbana.

Kitching, I. J., Forey, P. L., Humphries, C. J., and Williams, D. J. (1998). *Cladistics: The Theory and Practice of Parsimony Analysis,* 2nd ed. Systematics Association Publication 11. Oxford University Press, Oxford, United Kingdom.

Maddison, W. P., Donaghue, M. J., and Maddison, D. R. (1984). Outgroup analysis and parsimony. *Systematic Zool.* **33**, 83–103.

Nelson, G. J. (1973). The higher-level phylogeny of the vertebrates. *Systematic Zool.* **22**, 87–91.

Nelson, G. J. (1978). Ontogeny, phylogeny, paleontology, and the biogenetic law. *Systematic Zool.* **27**, 324–345.

Nixon, K. C., and Carpenter, J. M. (1996). On consensus, collapsibility, and clade concordance. *Cladistics* **12**, 305–321.

Watrous, L. E., and Wheeler, Q. D. (1981). The outgroup comparison method of character analysis. *Systematic Zool.* **30**, 1–11.

CLADOGENESIS

Christopher J. Humphries
The Natural History Museum, London

I. Diversification—Patterns and process
II. Anagenesis, cladogenesis, and stasis
III. Clades, grades, fossils, and ancestors
IV. Patterns of life
V. Rates of evolution
VI. Conclusions

GLOSSARY

adaptation Any genetically controlled characteristic that increases an organism's fitness, usually by ensuring the organism to survive and reproduce in the environment it inhabits.

adaptive radiation The evolution of one or a few forms into many different species that occupy different habitats within new geographical areas or habitats.

anagenesis A pattern of evolutionary change involving the transformation of an entire population, sometimes to a state different enough from the ancestral population to justify renaming it as a separate species; also called phyletic transformation of an unbranched lineage of organisms which changes to such an extent as to be justifiably called new species or taxa. This is considered "biological improvement" (sensu Huxley, 1958), covering all types of change from detailed adaptation to generalized organizational advance.

clade The branch between the nodes on a cladogram or phylogenetic tree; the unit of cladogenesis.

cladogenesis branching evolution; the origin of diverging new forms from an ancestral lineage; the form of divergence and phyletic splitting, including speciation; adaptive radiation of species to major divergence of families and phyla. Evidence for cladogenesis is from those fossil groups that increase in the number of recognizable taxa over time.

genotype The genetic constitution of individuals, species and taxa.

grade The product and "unit" of anagenesis; paraphyletic group.

Linnean system/hierarchy The hierarchical arrangement of inclusive categorical ranks in which, apart from the lowest rank, subordinates members of all lower ranks; it contrasts with exclusive classifications (e.g., *scala naturae*).

monophyly (monophyletic group) A natural group that includes a most recent common ancestor and all of its descendants.

natural classification Systems of classification portraying as accurately as possible the entire pattern of life and its relationships.

ontogeny The developmental history of an organism from egg to adult. Includes embryogenesis that describes the generative phase and the allometric phase of growth and maturity.

paraphyly A group that includes a most recent common ancestor and only some, not all, of its descendants.

phenotype The characteristics of an individual and all its parts as an interaction between the genotype and the environment.

phylesis Line(s) of descent between ancestors and descendants.

phylogeny The inferred lines of descent showing genealogical relationships of organisms; often used to describe nested series of monophyletic taxa both in cladograms and phylogenetic trees.

polyphyly (polyphyletic group) A group that does not include the most recent common ancestor of all of its members.

punctuated equilibrium The morphological stasis of fossil species over long periods of time punctuated by seemingly instantaneous speciation and morphological change.

Scala naturae (Great Chain of Being) One of the most pervasive ideas in western thought derived from two concepts Plato's principle of plenitude and the linear series of Aristotle and Plotinus (Panchen, 1992); in plain English, the belief in a linear progression from the simplest forms of life to the most perfect.

stasis (stasigenesis) The persistence of organisms through long geological periods of time that greatly resemble their fossil forebears. The "process" is used to recognize delimitable anagenetic units, or "grades."

synapomorphy Shared derived characters; homologies of monophyletic groups.

tree of porphyry A tree classification constructed as a dichotomous key that at any rank divides on the basis of differential characters to give two taxa at the rank below. The net result is a comblike branching diagram intended to classify individuals within a general scheme.

THE TERMS ANAGENESIS (divergent evolution), cladogenesis (branching evolution or diversification), stasis (stasigenesis; constancy), grades, and clades are all associated with mechanisms and patterns of phyletic evolution. Anagenesis is evolutionary change in lineages through time, cladogenesis is diversification of clades by branching of lineages through time (Rensch 1959), and stasigenesis (stasis) refers to persistence of lineages through time (Huxley, 1958). The distinctions came about through the belief that in evolution there were three main processes at work that led to "biological improvement," to "diversification," and to "persistence." Extinction, although seen by some to be of importance in evolution, is perceived as a failure to persist and is now treated as an important part of diversity assessments of life on earth. The concepts of phyletic evolution came to prominence in the neo-Darwinian synthesis of genetics and evolution, but their meanings changed when viewed through the lenses of different systematic perspectives. Grades, for example, once seen as extremely important with respect to the pattern of evolution in the *scala naturae*, are now seen as artifacts, described as paraphyletic and polyphyletic groups in the literature of cladistics where the emphasis lies in the analysis of clades and the recognition of sister groups. Furthermore, the processes of anagenesis and cladogenesis refer either to microevolutionary processes (speciation) or to patterns and rates of evolution through geological time calculated as rates of morphological, taxic, and molecular change.

I. DIVERSIFICATION—PATTERNS AND PROCESS

The centerpiece of evolutionary theory is that the environment varies with time and from place to place. Heritable variations that suit a particular environment are selected *in situ*, and so with time populations diverge and differentiate as each becomes adapted to its own conditions (Patterson, 1999). Darwin (1859) considered that such adaptive change led to diversity of form, with features thought to result from the process of diversification. Cetaceans and sea grasses, for example, are thought to be marine-adapted mammals and flowering plants, respectively, with land-borne ancestors. With the impact of tremendous changes in systematic thinking and population biology, Darwin's theory of organic evolution is now considered by most people to be the only worthy explanation of the diversity and form of life on earth. Darwin's theory has two quite different and distinct aspects. Evolution by common descent is used to describe the patterns of changes in diversity through time, and evolution by natural selection describes the processes leading to diversity as a measure of change over time.

The language of micro- and macroevolution emerged at a time when fossils were considered as of cardinal importance for understanding the patterns and processes ofvolution, when evolutionary theory, which was making great strides by amalgamating with genetics and evolutionary systematics, had yet still to mature. Evolution is the key. However, the greatest strides over the past 30 years have been made in systematics, where fossils have been analyzed together with recent taxa.

The relationships have changed, and systematics is the key. Progress in evolutionary theory has largely been in mechanisms. The concepts of cladogenesis and anagenesis have consequently changed. For example, as interpreted in the light of phylogenetic systematics (or cladistics) grades pertain to polyphyly and paraphyly, and clades to monophyly. The origin of species or higher taxa by anagenesis, cladogenesis, or phylogenesis, and persistence through stasis, is associated more with process biology in microevolution (see speciation processes) and particularly the assessment of patterns and rate of macroevolution.

A. The Synthetic Theory

The synthetic theory of natural selection (often described as neo-Darwinism) incorporates the original theory of natural selection of Darwin and Wallace (Darwinism) but also incorporates models of population genetics and the underlying bases of morphological and genetic heterogeneity within organisms. Classic genetic studies on model organisms such as *Drosophila, Cepaea*, and mimetic butterflies (see Panchen, 1992) have largely been concerned with genetic variation on different alleles in heterogeneous populations and the consideration of adaptive significance, which led to the rather stifling view that all change is adaptive. This has led to the idea that natural selection is essential for adaptive, anagenetic change and is a necessary precondition for cladogenesis (speciation). The processes of anagenesis, cladogenesis, and stasis are used to describe changes (or the lack thereof) in phenotypes and genotypes through changes in the environment both in space and time and by mutations and heritable variations within phylogenetic lineages. Patterns of diversity are described in phylogenetic schemes or genealogical trees through systematic analysis.

B. Speciation

As Huxley (1958) noted, Darwin (1859) recognized three processes at work “leading respectively to biological improvement, to diversification, and to persistence.” Improvement covers adaptation to specialized lifestyles, through phyletic transformation of lineages, resulting in the derivation of descendent species directly from ancestors. These are recognized as “transformational” or “evolutionary” species. New species arise as recognizable transformations within a lineage, by-products of the genetic properties of the lineage, and are thus recognized as arbitrarily divided segments of that lineage. Diversification covers all patterns and processes from micromorphological changes and subspeciation (microevolution), adaptive radiation (cladogenesis with adaptation), and macroevolution, the division into major phyla and higher taxa. Thus, the processes of divergence include lineage splitting, speciation, and adaptive radiation where genetic isolating mechanisms result in two or more descendants derived from a common ancestor. Persistence refers mostly to the idea that organisms of great geological age survive unchanged as living fossils into contemporary biotas. Little or no detectable change takes place through time such that the modern organisms resemble fossils from early geological periods. Examples include animals, such as *Latimeria* (Coelacanth), *Nautilus*, and *Sphenodon*, and plants, such as *Ginkgo* (Maidenhair tree), *Wollemia* (Wollemi Pine or “dinosaur tree”), and the *Metasequoia* (dawn redwood).

II. ANAGENESIS, CLADOGENESIS, AND STASIS

New species can either become modified through gradual change in an entire lineage in response to a changing environment or can emerge through diversification into two or more species through formation of internal and external isolating mechanisms. Anagenesis is gradual change in an entire lineage is (Fig. 1). Division into two or more species is termed cladogenesis (Fig. 2). There are many modes of population differentiation and considerable debate as to their roles in speciation. Likewise the mechanisms for anagenesis and cladogenesis (speciation) are many and varied. Nevertheless, they all include some form of population differentiation, either by gradual change of genetic differences and natural selection (phyletic gradualism), or by abrupt punctational changes, involving chromosome inversions or translocations and rapid isolation between populations,

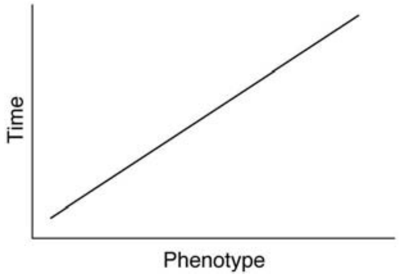


FIGURE 1 Anagenesis. The transformation of a species or taxon on an unbranched lineage of organisms. Divergence has occurred to such an extent that it is justifiably called a new species or taxon.

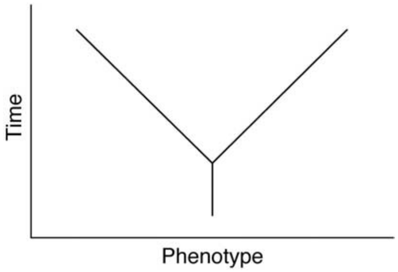


FIGURE 2 Cladogenesis. The transformation of a species or taxon into two (or more) species (or taxa) by branching along a lineage.

or by historical accidents, such as allopatric speciation by vicariance and isolation.

There is considerable controversy among the theories of diversification and how evolution proceeds. The Darwinian hypothesis is phyletic gradualism, whereby the same microevolutionary processes that lead to population differentiation cause ever-increasing divergence between populations (Fig. 3). Eventually, sufficient divergence has occurred for differentiation at species level to be recognized. Differentiation continues at a steady rate and new species originate by slow, gradual changes of ancestral species. The neo-Darwinian perspective is that evolutionary transformation takes place within species, or lineages, and the branching process of cladogenesis accounts for diversification but relatively small amounts of evolutionary change.

An alternative theory of evolutionary rates and speciation, punctuated equilibrium, was proposed by Eldredge and Gould (1972; Fig. 4). Evidence from paleontology on well-preserved fossils indicate long periods of stasis, where "species" remain relatively little changed over long periods of time. At other times there appears to have been rapid evolution and great morphological differentiation. Eldredge and Gould (1972) thus concluded that rates of evolution were not constant over time. They hypothesized that little evolutionary change occurs within species and that genetic changes within populations do not account for different species. Instead, the events of speciation account for evolutionary change, and short periods of "punctuation" were interspersed with little of no evolutionary divergence.

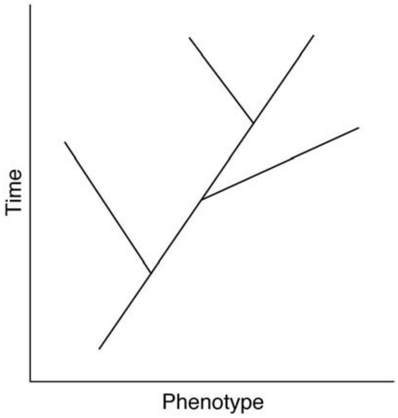


FIGURE 3 Phyletic gradualism. A fairly constant rate of change through time.

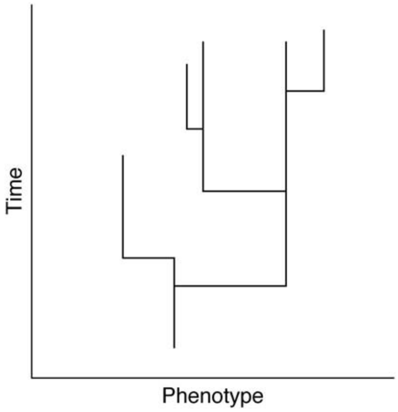


FIGURE 4 Punctuated equilibrium. Morphological stasis of fossil taxa over long periods apparently punctuated with occasional instantaneous change.

Phyletic gradualism and punctuated equilibrium represent the opposite extremes of a continuum. Recent research, especially on rates of molecular change, would suggest that evolutionary rates are clocklike, or at least "clocky," caused by periods of slow rates of phylesis and rapid periods of cladogenesis. This is reflected in research of the 1980s and 1990s, which has concentrated on uncovering the patterns of divergence through systematic analysis (particularly cladistics). Sustained application of cladistic methods has determined nature's hierarchy and techniques such as maximum likelihood have been used to assess rates of change amongst phylogenetic trees.

III. CLADES, GRADES, FOSSILS, AND ANCESTORS

Systematic classifications are represented in branching diagrams, variously know as trees, phenograms, dendrograms, and cladograms depending on which systematic philosophy is being used. For most practicing biologists of the 20th century, modern systematics dates back to Linnaeus. He recognized that nature's hierarchy could be arranged into a series of named categories.

His system of categories, although still in use today, is somewhat arbitrary, based on the predefined levels of *Regnum* (Kingdom), *Classis, Ordo, Genus*, and *Species*, to which Phylum and family were later added (Panchen, 1992). All the members of the same level were given the same rank. As distinct from exclusive systems, such as military ranks, biological ranks are all inclusive and genuinely hierarchical, whether viewed agglomeratively from the bottom or divisively from the top, such that eventually all taxa belong to one group, life itself. The hierarchy is divergent, so that a taxon of, say, a specific rank, belongs to a taxon of higher rank and the taxon of this rank belongs to a taxon of even higher rank, and so on. The hierarchy is not necessarily symmetrical, because taxa of high rank can have everything from one isolated member to a great number of members. Elements of Linnean system are still in use today, especially nomenclature, but its avowedly artificial category hierarchy, expressed as a preordained ranking system, has become replaced by a whole succession of "natural" classification systems portraying as accurately as possible the entire pattern of life and its relationships.

A. Natural Classification

Almost a century of debate has questioned the meaning of "natural" classification systems but by the time that the *Origin* was published (Darwin, 1859), the form of representation had become a branching diagram of hierarchical clusters. To his credit, Darwin made the tree metaphor explicit in what was the only diagram published in the *Origin* (Fig. 5). This was not a diagram produced by clear-cut rules but a tree rooted in the past—a genealogical tree to denote phylogeny in a vertical direction and classification of relative distances of present-day organisms in the branching hierarchy. One of the most important issues that came from Darwin's

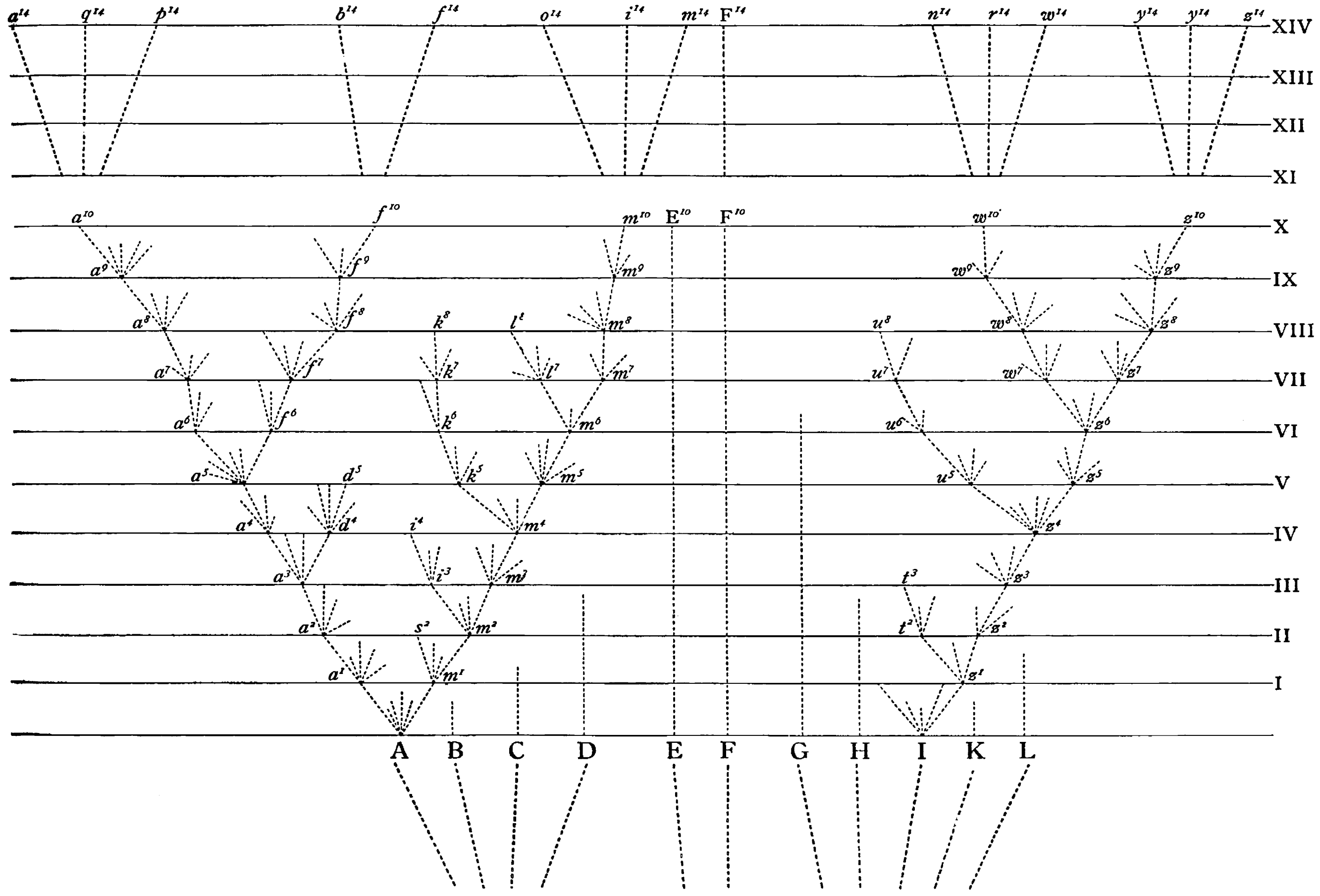


FIGURE 5 Darwin's (1859) branching diagram from the *Origin*. Note that A–L represent the species of a genus. The small letters show degrees of divergence through time. The horizontal lines represent 1000 or more generations. The diverging lineages are shown with extinct populations at most generations in every lineage.

evolution by common descent was that phylogeny is the theory that underpins natural classification today.

Fairly soon after the *Origin* appeared, "forests" of genealogical trees began to appear in the European literature. Perhaps the most memorable rendition of all schemes ever published is that of Haeckel (1874) who used literally the European oak (*Quercus robur*) to synthesize his interpretation of phylogeny and classification and the origin of man into a natural classification using Darwin's tree metaphor (Fig. 6). Apart from its shape, the named parts of the tree embody two quite separate concepts. The shape is simple enough in that it depicts a multifurcating inclusive genealogy with the names of the branches indicating monophyletic groups (groups derived from a common ancestor). However, the names on the vertical axis, horizontal lines or "grades" of evolutionary achievement, embody a watered down version of the *scala naturae*, a belief in the progression from the lowliest simplest forms of life (Monera) to the most "perfect" (human) through the series protozoa—metazoa—evertebrata—vertebrata—mammalia. The distinction was based on Haeckel's twin concepts of the evolution of the foetus, ontogeny, and recapitulation in the evolution of the stem (phylogenesis). Arising out of former was the phylogenetic interpretation of the *Tree of Porphyry* as a step series (Stufenreihe) and distinguished from the ancestor series (Ahnenreihe) in palaeobiology (see Panchen, 1992).

By the turn of the 20th century, branching diagrams were largely depicted as phylogenetic trees to represent "Ahnenreihe" as "grades" in ancestor-descendant series and "Stufenreihe" to represent evolution by cladogenesis into a succession of adaptive zones (Panchen, 1992; Fig. 7). There have been endless variations of this theme in comparative biology and different emphases placed on grades and clades. For example, Bessey's cactus (1915), deliberately emphasized grades and even interpreted extant supraspecific taxa, Ranales (Fig. 8a), as giving rise to other supraspecific taxa, Rosales (Fig. 8b), which in turn gave rise to Myrtales (Fig. 8c). At the other extreme, Janchen's (1932) interpretation of Wettstein's classification clearly emphasizes clades with grades added almost as an afterthought to indicate the distribution of floral evolution through gymnospermae/angiospermae, apetalae/sympetalae, and monocotoledony/dicotoledony (Fig. 9).

B. Grades and Clades in Evolutionary Systematics

These days we know that Bessey's approach is untenable—one supra-specific taxon cannot give rise to another. The Ranales cannot be the ancestor of the Rosales, and even if we did consider them to have an ancestor-descendant relationship then the Ranales and Rosales would have to be classified in the same group rather than as depicted in Figure 8. Nevertheless, grades and clades are still established practice and appear in textbooks even today. The 20th century has seen an enormous output of different phylogenetic schemes and the persistence of grades has pervaded the methods of "evolutionary systematics." The idea that grades are important in systematics so as to express patterns of anagenesis is the hardest notion to expunge. Grades are characteristic of classifications involving "satellite" groups considered sufficiently divergent to separate them at a high rank from their nearest relatives. The more "primitive" taxa are similarly rendered as a grade, a paraphyletic group in the language of phylogenetic systematics or cladistics.

To illustrate, perhaps the two most famous examples include human (*Homo sapiens*) considered superior to its primate relatives, and birds (Aves) considered greatly diverged from relatives in the class Reptilia (Fig. 10; see Mayr, 1974; Panchen 1992). The net effect of such schemes is that the phylogenetic reconstruction, based on character analysis, differs greatly from the written classification. Figure 10 illustrates this result in both examples. Thus for the diapsid reptiles the phylogenetic diagram shows ancestor A ancestral to two groups B (Diapsid reptiles) and C + D (crocodiles + birds). The classification differs in that group D (Aves) shows such considerable divergence along the horizontal character axis (the anagenetic component), that to express the information in a classification puts B + C together (diapsid reptiles + crocodiles) and D is kept separate (Aves). A similar story can be portrayed for orangutan + gorillas/chimps and *Homo sapiens*. The phylogenetic diagram shows C + D (gorillas/chimps and humans) as sister group to B (orangutan) but the classification puts B and C together and humans in a group D of their own. Mayr (1974) claimed that the anagenetic component should carry more weight because of many genes shared in common among birds and not found in the nearest relatives and because of the unique behavioral characters in man. Thus, groups represented by D are placed in higher grades and groups represented by B + C in lower grades.

C. Paraphyly and Monophyly in Cladistics

Cladistics demonstrates that grades are impossible to characterize as they do not express relationships between organisms. There is no way of knowing when

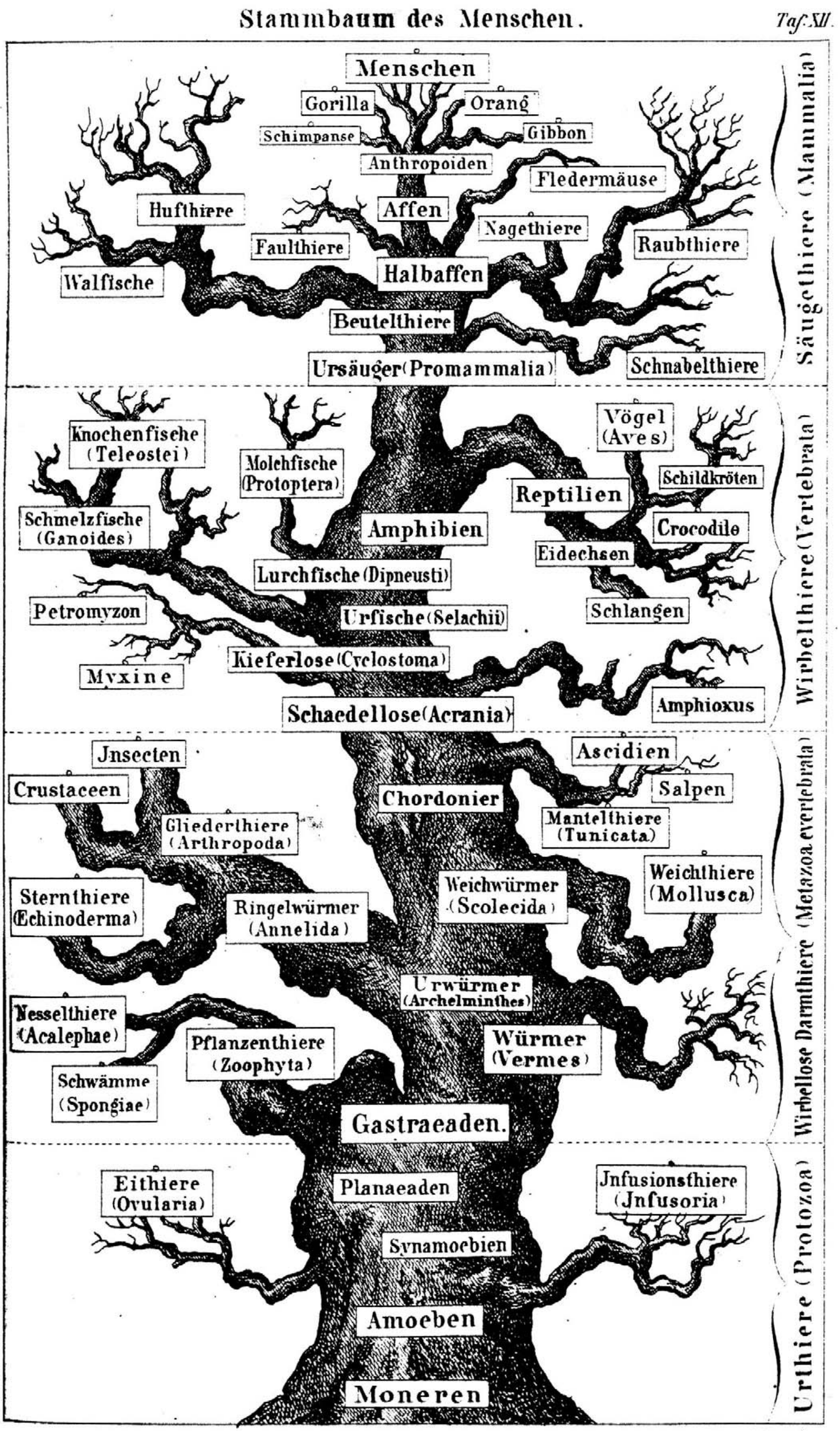


FIGURE 6 Haeckel's (1874) "literal tree" (European oak) depicting human ancestry as interpreted by Darwin's theory of evolution by common descent. The bifurcating branches indicate groups determined on embryology and the vertical axis (conflated with time) represent grades.

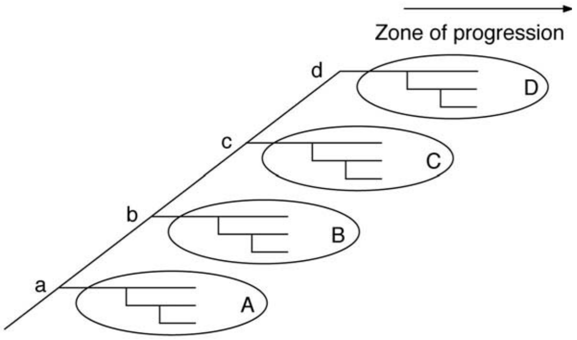


FIGURE 7 Abel's concept of Stufenreihe (clades A–D) and Ahnenreihe (grades a–d) (After Panchen, 1992; Fig. 3.6, p. 55). See text for explanation.

one grade starts and another stops. How many characters along a branch are required before it is possible to draw a line and say that one side represents a lower grade and the other side a higher grade? Related to this question is how many taxa belong to a grade group?

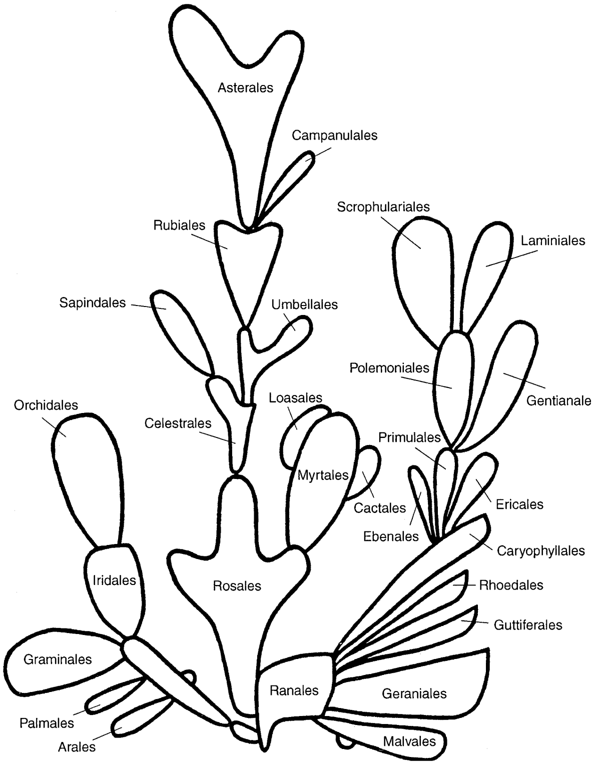


FIGURE 8 "Bessey's cactus" (Bessey, 1915, Fig. 1, p. 118). Diagram to show the relationships of orders. The size of each order is represented by the shape and dimensions of each component of the diagram. For an explanation of a, b, and c, see text.

Hennig (1996) described a method to implement evolution by common descent by reconstructing phylogenies based on assessments of speciation (cladogenesis) and transformation of characters (now known as cladistics; see Methods of Systematics, and Cladistics). His most important contribution was to offer a precise definition of relationship and a technique for those relations to be discovered. A minimum of three taxa is necessary to express a relationship. For example, in Figure 11a taxa B and C are more closely related to each other, than either is to A, because they share a common ancestor not shared by A. Cladistic analysis finds monophyletic groups on the basis of uniquely derived, shared characters (synapomophies).

Hennig showed that monophyletic groups are "natural groups," that branching nodes expressed relations, and that synapomorphies are the only measurable quantities for determining pattern. Hence neither horizontal nor vertical branches (anagenesis) are meaningful for expressing relationships on cladograms or for the determination of groups. Vertical branches say nothing about time and the relative nesting of nodes on the cladogram provide only relative rather than absolute estimates of ordinal time. Relationships could just as easily be represented by nested sets or Venn diagrams (Fig. 11b).

Cladograms are synapomorphy schemes, induced from the most parsimonious distribution of characters to show sister-group relationships. In Figure 12a, taxa B and C represent one sister group nested in larger sister group, A and B + C. Cladograms are different from phylogenetic trees because they rely entirely on empirical data, taxa, and characters. They express only the general branching pattern of life because that is all that is available from analysis of taxa and form. The strongest support for this idea is that many phylogenetic trees can be hypothesized for the same cladogram irrespective of branch length (Fig. 12 b–e). Consequently, anagenesis, ancestors, and ancestor-descendant relationships are not directly available from character analysis but require models of one kind or another to arrive at answers to questions of rates of divergence. Cladograms are different from phylograms or phylogenetic trees because they are isomorphic with the classification. Cladograms are consistent with name hierarchy and they can be recovered from written classifications. On the basis of this property, Hennig (1966) justifiably claimed that phylogenetic systematics provided the only truly general reference system consistent with the theory of evolution by common descent.

The task of systematics for Hennig was to understand natural relationships (monophyly, monophyletic groups) and rid the general reference system of polyphyly and paraphyly. Although there has been consid-

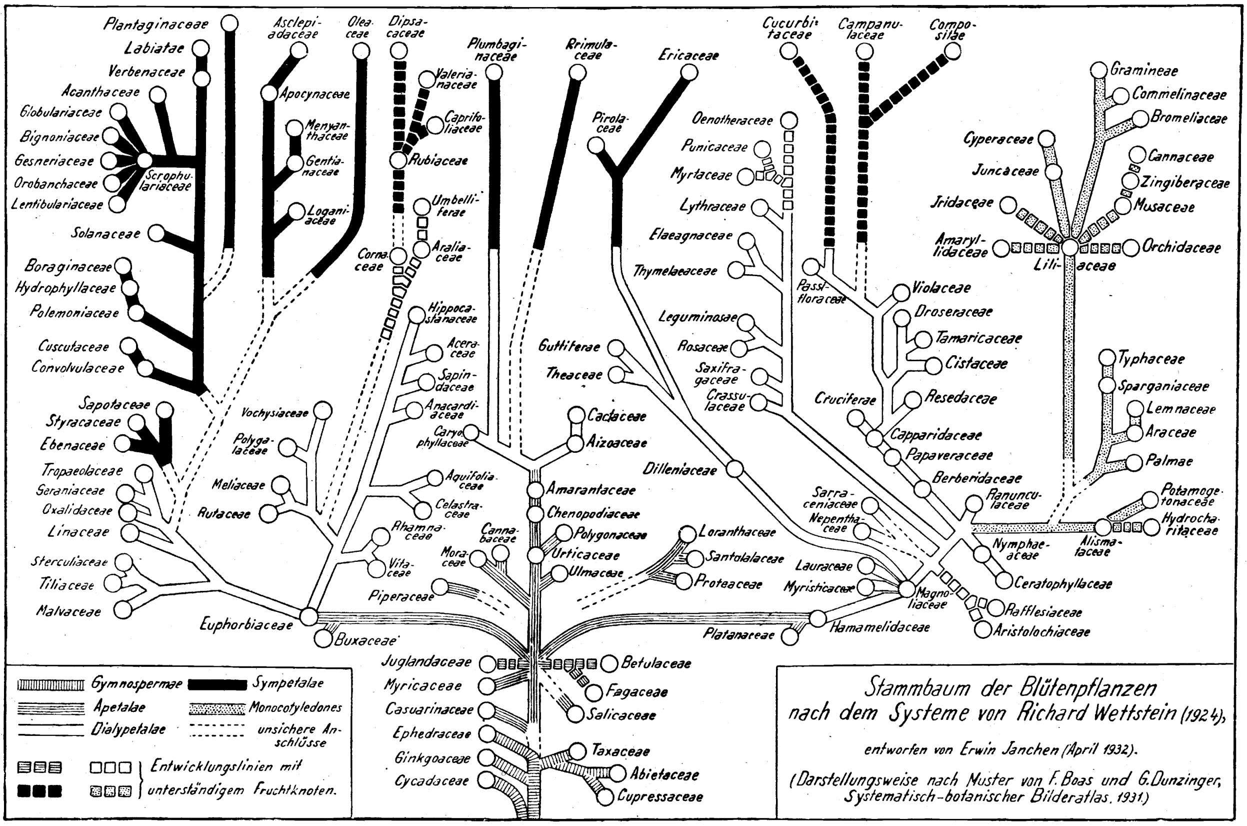


FIGURE 9 Grades and clades, Janchen's (1932) phylogeny. The phylogeny of the seed plants is shown as a branching diagram. Grades are depicted by irregular, thick black lines crossing the branches of the phylogeny.

erable proliferation in the methods and sources of information in systematics, the main effect has been to put intense effort into cladogenesis by the discovery of clades or monophyletic groups. For the past 25 years or so systematics has concentrated on determining the pattern of life from its earliest beginnings to the highest nodes on the tree, especially as a result of massive strides in molecular biology. Programs in pattern analysis range from the minutiae of phylogeography within species and populations to the discovery of monophyletic clades throughout the entire history of life. On the process side have been intense efforts to discern the rates of macroevolution by calibrating what is known about fossil history and morphological evolution with what is known about base substitution rates in ubiquitous molecules.

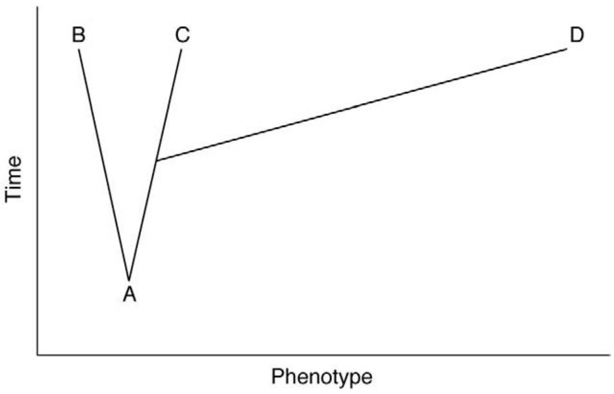


FIGURE 10 Grades and satellite groups. A = ancestor; B = diapsid reptiles, or orangutan; C = crocodiles, or gorillas/chimps; D = birds, or humans. See text for explanation.

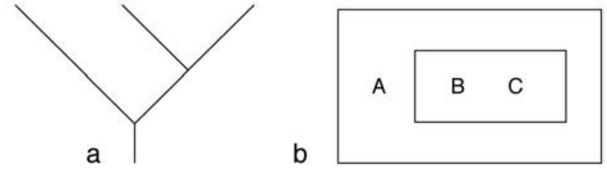


FIGURE 11 Definition of relationship in (a) a cladogram and (b) a Venn diagram.

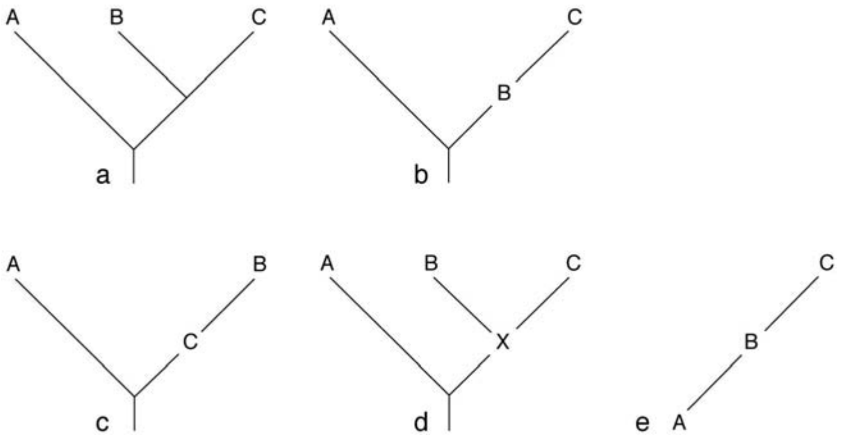


FIGURE 12 Cladograms and trees: (a) cladogram, (b–e) a selection of phylogenetic trees consistent with the cladogram in (a).

IV. PATTERNS OF LIFE

The emphasis over the past 30 years has switched from worrying so much about whether one species or taxon gave rise to another species or taxon to emphasizing phylogenetic trees and cladistic patterns. The emphasis has switched entirely from ancestor-descendant stories to expressions of sister-group relationships. The emphasis on the overall morphology and anatomy of the phenotype has comparably given way to the use of molecular characters. Furthermore, the discovery that sequences of amino acids of protein chains and nucleic acids contain detailed sources of information has given impetus to the discovery of family trees across the entire hierarchy of life. By comparing homologous amino acid or gene sequences it has been possible to compile family trees using computerized cladistic analyses with thousands of taxa in one calculation. Through a massive empirical enterprise, involving literally thousands of published investigations, the stage has been reached when comparisons between bacteria, flowering plants, and mammals in the same analysis are a reality. The potential for observing the patterns on both a grand scale and at species level is dominating comparative biology in all disciplines. The output of data and cladograms is so vast and the information coming from a variety of museums and laboratories around the world that dedicated web sites are being built to makes sense of it all (see, for example, The Tree of Life web page at http://phylogeny.arizona.edu/tree/phylogeny.html).

The overriding message from all of this empiricism is that large-scale differences between humans, mice, elephants, and oak trees have the same sort of causes as closely related species (Patterson, 1999). It appears that at all scales of investigation the differences between organisms related by common ancestry owe their differences to divergence, the accumulation of mutations over great periods of time. The evidence for this homology comes from the fact that organisms share homologous features at many levels, molecular, structural, and physiological. At face value it would seem that gradual phylesis would explain changes in both micro- and macroevolutionary events. Several evolutionists have different opinions about this. They can accept point mutations, gene duplications, and chromosome mutations to explain the smaller divergences in birds and butterflies but require other mechanisms to suggest the massive differences between fishes and land organisms or the original appearance of crustacea and green plants. The main reason for requiring macromutations for macroevolution is that many features of major groups cannot be imagined to have come into being by small adaptive changes. The point often made is the existence of complex organs. Familiar questions are of what use is half an eye or half a lung? The answers are not readily apparent because the evidence is unclear. The intermediate steps cannot be imagined. By improving the accuracy of reconstruction of cladograms for the whole of life the potential exists for assessing whether the story is one of gradual phylesis or not.

V. RATES OF EVOLUTION

Rates of evolution have always been of interest to systematists and evolutionists alike, and one of the popular myths is that the fossil record is the only direct evidence to track the course of evolution. During the history of paleontology many arguments have been constructed to link the fossil record to evolution because we know that fossils are the remains of once living organisms and they predate the rock sediments in which they lie (Panchen, 1992). Furthermore, the idea that the stratigraphic record encapsulates the historical record, and that most fossils are extinct but related to modern groups that extend a little way back in time, is the underlying justification for schemes to correlate the pattern of fossil record with evolution through time. It has long been considered possible to calculate different rates for a number of reasons. "Living fossils" that have survived from the geological past to the present-day show that most organisms in the fossil record were once living but are now extinct. Similarly, "Lazarus" taxa, found in an early geological period, which then disappear for millions of years to reappear in more recent periods, were considered to be a good line of evidence to support phylogenetic reconstructions with absolute, rather than relative time, intervals. The fact that many modern groups are modified older groups and can be traced back in time has reinforced evidence for the transmutation of fossil organisms into modern forms and hence the belief that rates of evolution can be calculated.

A. Morphological Evolution

Perhaps the most cited examples of evolution between fossils and modern taxa are the so-called missing links. Fossils, such as *Archaeopteryx*, were considered to show links between different major groups. The apparent intermediate nature of *Archaeopteryx* by having teeth, as well as derived features of feathers and wishbones, as found in all birds, was used to suggest that birds are descendants of dinosaur-like creatures. No exact sequence of fossils or ancestors could ever be found and it has been known for a long time that no particular group of fossils could give rise to a particular group of modern taxa. The search for ancestors has gradually switched from reconstructing history sequences of grades in a *scala naturae* to the modern approach of comparing dated fossils in stratigraphic sequences with well-corroborated cladograms (Smith, 1994).

There have been many reconstructions of rates of evolution through geological time in fossil and modern taxa of plants, vertebrates, and invertebrates (see Smith, 1994). The oft cited examples are mammals, particularly horses, from the Eocene to the present, because of their good fossil record (see Panchen, 1992; Simpson, 1951, 1953). In the case of horses, the technique was to see how well morphoclines of characters from the most primitive characters in the oldest taxa gradually changed into the derived characters of modern taxa. These were then examined to see how well these matched the chronocline (the age of fossils as determined by their appearance in the stratigraphic sequence). Simpson's (1951) pioneering work traced lines of descent from the fossil *Eohippus* to *Equus* by creating a linked time series, or *scala naturae*, to measure the rates of morphological changes in skull, forelimb, and upper molar characters in a time period of 60 million years. The approach used was to literally measure the amount of change over a generalized phylogenetic tree after dates were assigned to the horse genera in the sequence (Fig. 13). Despite the great effort expended by Simpson, his results give only general approximations on the rates of divergence and the absolute dates

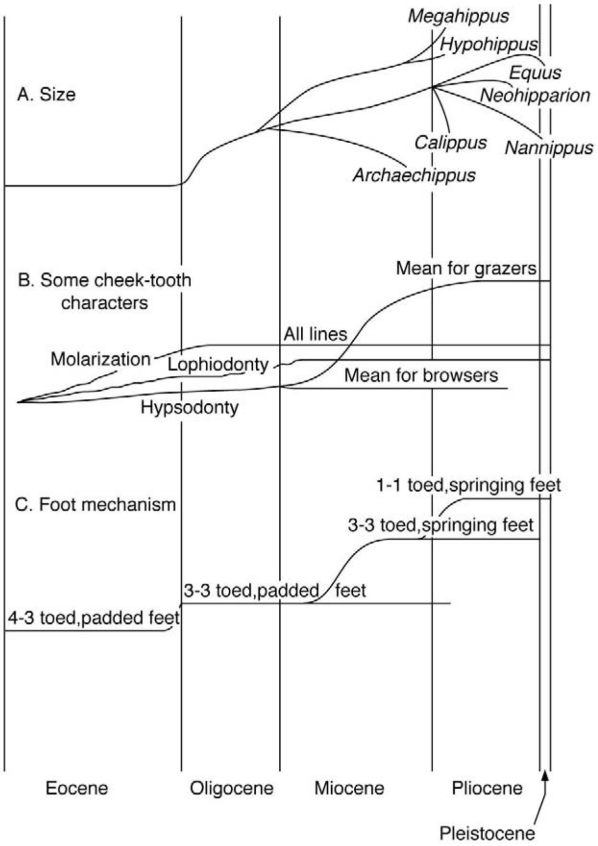


FIGURE 13 Evolution of fossil horses. (A) represents a phylogram of increasing size, (B) changes in cheek teeth characters, (C) represents number of toes and foot mechanisms. After Simpson (1953, Fig. 34, p. 265).

are hard to accept because of a lack of a well-corroborated cladogram.

B. Rates of Molecular Evolution—Hemoglobin

The flood of molecular data accrued since the mid-1960s offers the prospect of more accurate assessments in the calculation of rates of evolution. Absolute rates of molecular evolution have been estimated from branching diagrams by calculating trees from various forms of sequence data and alignment of the cladograms to fossil or biogeographic events as fixed starting points. By calculating pairwise distances, the minimum ages of cladogenetic events across pairs of taxa can be estimated. These calculations assume that a "clock" is operating within the molecules by gradual and regular substitution of nucleotides or amino acids constantly over time and that the clock can be calibrated to external events (e.g., the age of a fossil), which then reflect accurately the initial cladogenetic events.

Early estimates in the 1960s were made on chains of amino acids and earliest occurrence of fossils to calculate rates of divergence in proteins. Typical results for alpha haemoglobins (Table I) comparing recent organisms with their fossil relatives suggest that sharks, coelocanths, and lungfishes have hardly changed over 350 million years or more but mammals and birds have transformed totally from their fishy ancestors (Patterson, 1999). Patterson showed these calculations indicate that hemoglobin in humans and elephants differ by about 18% as does that of the starling and ostrich. Also, humans and elephants diverged from the ancestral mammals at roughly the same time that ostrich and starling diverged from ancestral birds. In terms of amino acid differences, crocodiles differ from birds by about 33% and mammals by about 35%.

From the rough phylogenetic tree (Fig. 14), Patterson calculated from the distance data with a corrected time scale relative rates of amino acid substitution (Table I). Because there were more than 60% differences in the hemoglobin of sharks, as compared with humans, Patterson corrected for the error of multiple hits, because logically the real number of differences between the different taxa must be greater than those observed. The adjusted divergence times gave 140 myrs for the Platypus/mammal and crocodile/bird splits and about 70 myrs for the elephant/human and ostrich/starling divergence times.

It was from these kinds of results that questions of how natural selection could keep constant rates of divergence through time were asked. One such theory was the Red Queen hypothesis. It takes its name from *Alice in Wonderland* because every species has to keep on running just to keep in the same place. In other words, to compete and to survive all of nature's vicissitudes during the evolutionary changes from shark to human, and between any other pair of species, it is necessary for hemoglobin, and presumably all other features, to adapt at a constant rate (Patterson, 1999).

C. Rates of Molecular Evolution—DNA

With the advent of nuclear and plastid sequences in the late 1970s and a routine availability of sequences

TABLE I

Differences (Distances) in the Alpha Hemoglobin Chains of a Selection of Vertebrates

	Human	Elephant	Platypus	Ostrich	Starling	Crocodile	Lungfish	Coelacanth	Goldfish	Shark
Human	—	26	40	43	41	47	83	70	68	71
Elephant	18	—	45	45	48	50	84	72	63	74
Platypus	28	32	—	54	52	51	89	74	70	76
Ostrich	30	32	38	—	26	36	91	75	68	73
Starling	29	34	37	18	—	47	91	77	67	70
Crocodile	33	35	36	26	33	—	85	78	70	77
Lungfish	59	59	62	64	64	59	—	90	94	86
Coelacanth	49	51	52	53	54	55	63	—	83	78
Goldfish	48	44	49	48	47	49	66	58	—	88
Shark	50	52	54	52	50	55	63	55	62	—

The alpha chain contains 141 amino acids in most vertebrates, but it contains 142 in the goldfish and the lungfish has 143. Numbers to the upper right half are differences and those in the lower left are percentages. (From Patterson, 1999, Table 9.1.)

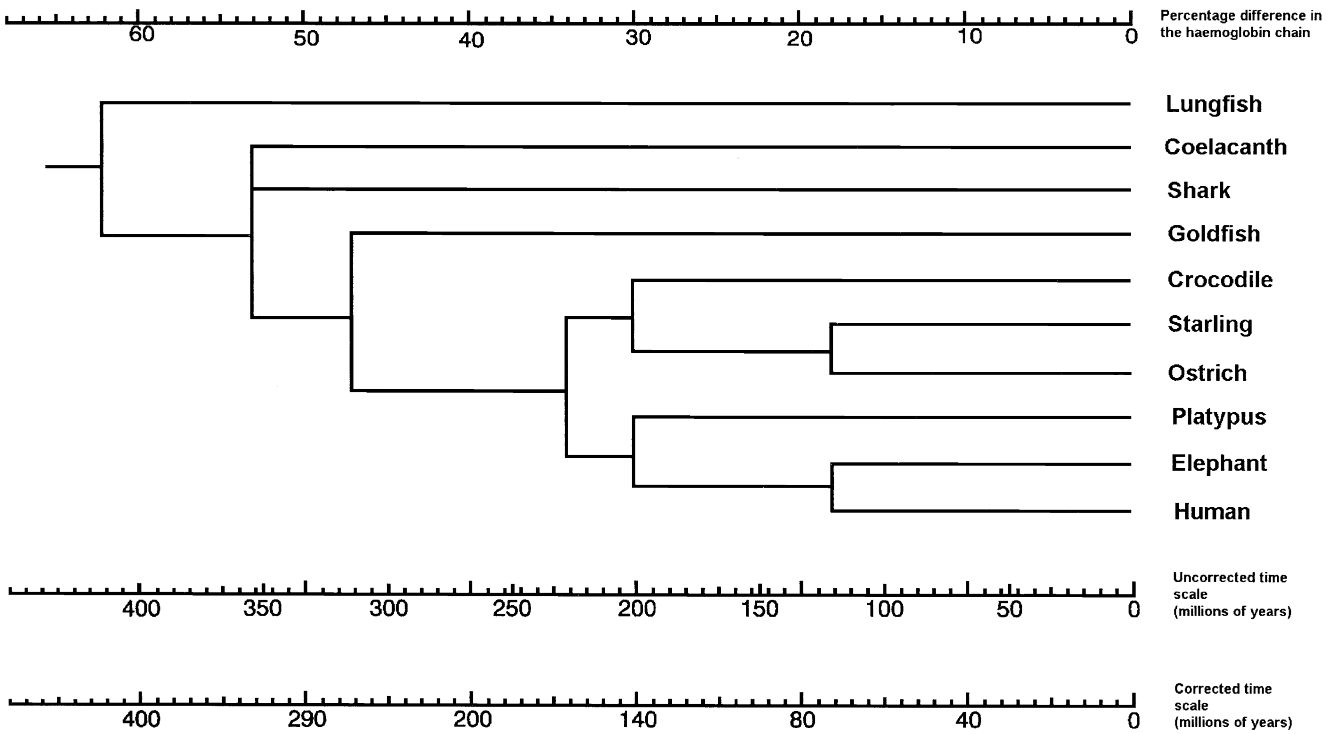


FIGURE 14 Hemoglobin family tree based on differences between the alpha hemoglobins in Table I. The two time scales are calibrated from the assumption (based on the earliest known fossils) that sharks, lungfishes, and coelocanths diverged about 400 mybp with 60% amino acid divergence calibrated at 400 myrs. (Reproduced from Patterson, 1999; Fig. 9.1.)

becoming available by the 1990s, it has become increasingly popular to infer divergence time estimates based on nucleotide sequences directly. Through a series of trials and errors it had become obvious by the mid-1990s that single point estimates of divergence times without any assessments of standard error were hardly worth calculating, and some greater degree of sophistication was needed (Waddell *et al.*, 1999). Recent calculations have attempted to improve results by incorporating differential substitution rates for molecular sequence change, a wider range of reliable calibration points based on fossils and other sources of information, and incorporation of standard error values to factors that can be quantified. Waddell *et al.* (1999) applied these techniques to estimate rates using maximum likelihood models and the origins of major lineages of mammals and birds using all available published mitochondrial DNA protein sequences (including unpublished ones for elephants and birds (Figs. 15 and 16).

For mammals (Fig. 15) the results show that the well-studied horse/rhinoceros split diverged about 55 million years before present. The splitting time among carnivores is confidently shown to be more than 50 million years, and splits within the placentals are relatively old, in the case of armadillo and other mammals 169 mybp but within 2 standard errors of other estimates. Waddell *et al.* reported also that the whale/cow divergence at 65 mybp may be much older than previously assumed. One of the interesting findings is that the sampled splits between the main groups of fereuungulates (carnivores, cetartiodactyls, perissodactyls, and pholidotes) took place prior to the critical Cretaceous/Tertiary boundary. The date of the vicariant separation of Africa and South America is coincident with the age of the cladogenetic event of the armadillo/elephant close relationship at around 122 mybp and could be a causal factor.

The data for birds (Fig. 16) was considered more controversial by Waddell *et al.* than for mammals. This was because the deeper calibration points normally used by other researchers between either crocodiles or mammals and birds were considered controversial in their assumptions that rates of evolution in birds had not changed much. They used the fossil anseriform, *Prebyornis*, and placed it with the ducks (geese and swans) using a constraint of anseriform monophyly to have

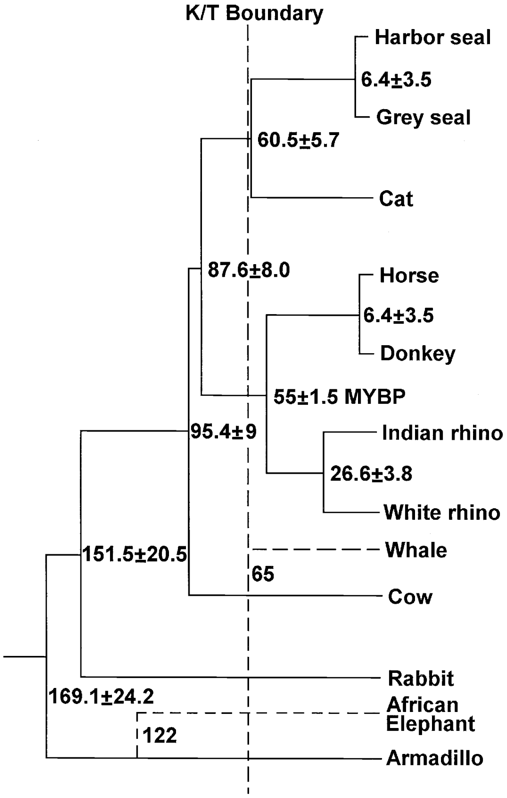


FIGURE 15 A clock-constrained maximum likelihood model mitochondrial DNA tree for mammals. The dates at each node represent millions of years before present (Reproduced from Waddell *et al.*, 1999, Fig. 1.)

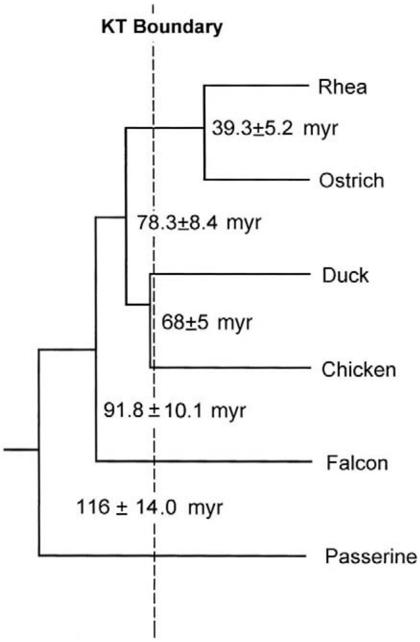


FIGURE 16 A clock-constrained maximum likelihood mitochondrial DNA tree for birds. The dates at each node represent millions of years before present (Reproduced from Waddell *et al.*, 1999, Fig. 2.)

at least one calibration point within birds. Using this information together with more detailed studies of anserifom relationships gave an anseriform stem lineage date of 58–78 (midpoint 68) mybp. The divergence times are indicated in Figure 16. The interesting findings are that the passerines seem to have split off from the rest of the birds at a much earlier date than normally believed and that the divergence time of ratites (rhea and ostrich) is much younger than generally believed.

IV. CONCLUSIONS

In an online *Encyclopaedia of Genetics*, Ernst Mayr recently considered cladogenesis and anagenesis as the two great phylogenetic processes of biology. He said that cladogenesis is the study of the origin and of the nature of the branching pattern of the phylogenetic tree. It deals with the various different methods by which the phylogenetic tree is reconstructed. Also, it includes the process of speciation, because every act of speciation adds a branch, no matter how short, to the phylogenetic tree. It is difficult to argue with such points of view because the definitions encapsulate so much of what we know and do in the modern synthesis. They are catchall statements with little or no precision in the contemporary scheme of things. The position taken here is that cladogenesis, and all of its related terminology, sits in some kind of neo-Darwinian time warp, which has become overtaken by huge developments in both systematics and evolutionary analysis. As Mayr points out, cladogenesis in a very particular sense refers to microevolution and speciation, both topics dealt with elsewhere. In macroevolution the interests are centered around the reconstruction of cladograms and in the discovery of monophyletic groups, topics considered elsewhere in this encyclopedia. In cladistics all effort is concentrated on the patterns of clade distribution and it is generally accepted that grades and branch lengths are of no value in phylogenetic classification but are of interest for determining rates of change. Cladogenesis, anagenesis, and stasis are general descriptors of studies of rates in evolution and extinction, as expressed here in relatively recent calculations of ages of taxa and deeper evolutionary splits. In conclusion, the importance of cladogenesis and related terms have become uncritical in the light of relatively recent studies in systematics; rather they are words in an arcane language largely consigned to the history books.

See Also the Following Articles

ADAPTATION • ADAPTIVE RADIATION • CLADISTICS • EVOLUTION, THEORY OF • FOSSIL RECORD • PHYLOGENY • SYSTEMATICS, OVERVIEW • TAXONOMY, METHODS OF

BIBLIOGRAPHY

Bessey, C. E. (1915). The phylogenetic taxonomy of flowering plants. *Annals Missouri Botanical Garden* **2**, 109–164.

Darwin, C. (1859). *The origin of species by means of natural selection*. London: Murray.

Eldredge, N., and Gould, S. J. (1972). Punctuated equilibria: An alternative model to phyletic gradualism (pp. 82–115). In T. J. M. Schopf (Ed.), *Models in paleobiology*. San Francisco: Freeman, Cooper & Co.

Haeckel, E. (1874). *Anthropogenie: Keimes- und Stammes-Geschicte des Menschen* (1st ed.). Leipzig: W. Engelmann.

Hennig, W. (1966). *Phylogenetic systematics*. Urbana: University of Illinois Press.

Huxley, J. S. (1958). Evolutionary processes and taxonomy with special reference to grades. Uppsala Univ. Årsskr. **6**: 21–39.

Janchen, E. (1932). Entwurf eines Stammbaumes der Blütenpflanzen nach Richard Wettstein. *Öster. Bot. Zeit.* **81**(3), 161–166.

Mayr, E. (1974). Cladistic analysis or cladistic classification. *Z. Zool. Syst. Evolforsch.* **12**, 94–128.

Panchen, A. L. (1992). *Classification and evolution, and the nature of biology*. Cambridge: Cambridge University Press.

Patterson, C. (1999). *Evolution* (2nd ed.). London: The Natural History Museum.

Rensch, B. (1959). *Evolution above the species level*. London: Methuen.

Simpson, G. G. (1951). *Horses*. New York: Oxford University Press.

Simpson, G. G. (1953). *The major features of evolution*. New York: Columbia University Press.

Smith, A. B. (1994). *Systematics and the fossil record; documenting evolutionary patterns*. Oxford: Blackwell Scientific Publications.

Waddell, P. J., Cao,Y., Hasegawa, M., and Mindell, D. (1999). Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. *Systematic Biology* **48**(1), 119–137.

CLIMATE CHANGE AND ECOLOGY, SYNERGISM OF[1]

Stephen H. Schneider* and Terry L. Root†
**Stanford University; †University of Michigan*

GLOSSARY

climatic models Mathematical descriptions of the flows of energy, momentum, and materials around the atmosphere and oceans, typically run on large computers to stimulate the earth's climate naturally and when disturbed by greenhouse gases or other so-called forcings or disturbances.

global change disturbances Alterations or disturbances to the natural conditions of the global atmosphere (e.g., greenhouse gases produced by human activities) or to enough regions (e.g., largescale deforestation or the wide-spread introduction of exotic, invasive species in many places throughout the world) to have global impacts.

global warming The additional heating of the earth's climate from the incremental injection of greenhouse gasses to the atmosphere from human activities such as deforestation or fossil fuel consumption.

greenhouse gasses Gasses in the atmosphere which selectively let more of the solar energy into and out of the atmosphere than they permit the transmission of long-wave infrared radiation. They contribute to the greenhouse effect, which traps radiant energy in the lower atmosphere and makes the earth warmer.

synergisms Interactions among many factors, which collectively may have a much larger or smaller effect than the sum of the effects of each factor acting independently.

CLIMATE HAS VARIED SUBSTANTIALLY on geological timescales, and ecological systems have responded with large shifts in vegetation and animal patterns, including extinctions. Changes of about 5°C over many thousands of years have occurred and triggered the ecological responses. Human activities which clear land or burn fossil fuels inject greenhouse gases such as carbon dioxide (CO_2) or methane (CH_4) into the atmosphere. There has been a 30% increase in CO_2 and a 150% increase in CH_4 since the industrial revolution. Climatic theory—supported by climatic models run on supercomputers, suggests that several degrees Celsius of global warming is possible—even likely—during the

---

[1] Adapted from Schneider and Root (1998).

next century. Ecologists are concerned that this could substantially rearrange the ranges and interactions of many species. However, because of human land uses such as agriculture, urban settlements, or roads, species no longer have a free range in responding (e.g., by migrating) to natural climatic shifts. The combination of effects—synergism—among climate changes, land-use disturbances, the introduction of exotic species and artificial chemicals may collectively impact on wildlife and terrestrial systems much more significantly than if each of these disturbances were simply considered separately. Therefore, many scientists and others have serious concerns about the need to conserve wildlife and vegetation in the face of such plausible global change disturbances.

I. SYNERGISMS OF CLIMATE CHANGE AND ECOLOGY

The earth's climate is vastly different now from what it was 100 million years ago when dinosaurs roamed the planet and tropical plants thrived closer to the poles. It is different from what it was 20,000 years ago when ice sheets covered much of the Northern Hemisphere. Although the earth's climate will surely continue to change, climatic changes in the distant past were driven by natural causes, such as variations in the earth's orbit or the carbon dioxide (CO_2) content of the atmosphere. Future climatic changes, however, will probably have another source as well—human activities. Humans cannot directly rival the power of natural forces driving the climate, for example, the immense energy input to the earth from the sun that powers the climate. We can, however, indirectly alter the natural flows of energy enough to create significant climatic changes. The best-known way people could inadvertently modify climate is by enhancing the natural capacity of the atmosphere to trap radiant heat near the earth's surface—the so-called greenhouse effect. This natural phenomenon allows solar energy to reach the earth's surface and warm the climate. Gases such as water vapor and CO_2, however, trap a much larger fraction of long wavelength radiant energy called terrestrial infrared radiation near the earth's surface. This causes the natural greenhouse effect to be responsible for approximately 33°C (60°F) of surface warming. Thus, seemingly small human-induced changes to the natural greenhouse effect are typically projected to result in a global warming of 1–5°C in the next century (Intergovernmental Panel on Climate Change, 1990, 1996a). This could result in an ecologically significant change, which is why climatic considerations are fundamental in the discussion of the status and trends of ecological conditions.

We may already be feeling the climatic effects of having polluted the atmosphere with gases such as CO_2. Many activities associated with human economic development have changed our physical and chemical environment in ways that modify natural resources. When these changes, such as burning fossil fuels that release CO_2 or using land for agriculture or urbanization that causes deforestation, become large enough, significant global (worldwide) changes are expected. Such modifications can disturb the natural flows of energy in Earth systems and thus can force climatic changes. These disturbances are also known as global change forcings. Quantitative evaluations of the potential effect of human activities in creating global change are needed. Such evaluations are also central to potential policy responses to mitigate global changes (Schneider, 1997; Intergovernmental Panel on Climate Change, 1996c).

A. Synergisms

One of the most potentially serious global change problems is the synergistic or combined effects of habitat fragmentation and climate change. People fragment natural habitats for farmland, settlements, mines, or other developmental activities. If climate changes, individual species of plants and animals will be forced to adjust if they can, as they have in the past. It seems unlikely that all of the migrating species that survived the Ice Age would be able to safely reach refuges after migrating across freeways, agricultural zones, industrial parks, military bases, and cities of the twenty-first century. An additional complication arises with the imposition of the direct effects of changes in CO_2, which can change terrestrial and marine primary productivity as well as alter the competitive relations among photosynthesizing organisms.

One representative instance of synergism is that of the Kirtland's warbler in northern Michigan; this species is restricted to a narrow area of jack pines that grow in sandy soil (Botkin *et al.*, 1991). Forest gap models of growth and decline of jack pines indicate that this species will move north with warming, but the Kirtland's warbler will not likely survive the transition. This bird nests on the ground under relatively young pines, and the soil to the north is not generally sandy enough to allow sufficient drainage for successful fledging of young (Cohn, 1989). Consequently, global warming could well doom the warbler to extinction in 30–60 years. This potential for extinction indicates how the already high rate of extinctions throughout the world would be exacerbated by climatic changes occurring more rapidly than species can adapt (Pimm, 1991; Peters and Lovejoy, 1992; Wilson, 1992).

The synergism question raises a controversial management problem of anticipating global change risk and responding by setting up interconnected nature reserves to ensure against some species becoming extinct in the event of climate changes. Alternatively, we could simply let the remnants of relatively immobile wildlife and natural plant communities remain in isolated reserves and parks as now exist. If we do opt for more environmental safeguards by interconnecting our parks, the question then becomes how do we interconnect the nature reserves. Priorities must be set and money made available for constructing natural corridors through which species can travel. For example, elevated sections of highways may be needed to allow for migration routes, similar to what was done for the caribou in the Arctic when the Alaskan pipeline was built. In order to examine such questions as these in scientific detail, it is first necessary to make a multi-disciplinary examination of the sub-components of the various aspects of climatology and ecology. We begin with a background discussion of climatic history, processes, modeling, and validation, as a prelude to focusing on ecological processes, which need to be examined in order to project possible synergisms among ecology and climate change.

II. CLIMATE HISTORY: WHAT HAS HAPPENED

Scientists can reconstruct the cyclical expansion and contraction of polar caps and other ice masses from ice core samples taken from Greenland and Antarctica. When snow falls on high, cold glaciers the air trapped between snow grains is eventually transformed into air bubbles as the snow is compressed into ice from the weight of subsequent accumulations. The ratio of two oxygen molecules with different molecular weights (O^{16} and O^{18} isotopes) is a proxy record for the temperature conditions that existed when the snow was deposited. From this, scientists have been able to determine that the ice buildup from 90,000 to 20,000 years ago was quite variable and was followed by a (geologically speaking) fairly rapid 10,000-year transition to the (current) climatically very stable Holocene Period. The Holocene is the 10,000-year interglacial period in which human civilization developed and modern plant and animal distributions evolved to their current states (Eddy and Oeschger, 1993). These ice cores also provide information on the presence of CO_2, an important greenhouse effect gas. Carbon dioxide was in much lower concentrations during cold periods than in interglacials (which is similar for the greenhouse gas methane, CH_4). This implies an amplifying effect, or a positive feedback, because less amounts of these gases during glacials means less trapped infrared radiative heat amplifying the cooling, and vice-versa during interglacials. The ice cores also show that concentrations of CO_2 and CH_4 and temperature were remarkably constant for about the past 10,000 years (before A.D. 1700), particularly when compared with the longer record. That relative constancy in chemical composition of the greenhouse gases held until the industrial age during the past two centuries.

The transition from Ice Age to the Holocene took 5,000 to 10,000 years, during which time the average global temperature increased 5–7°C and the sea level rose 100 m. Thus, we estimate that natural rates of warming on a sustained global basis are about 0.5–1°C per thousand years. Such changes were large enough to have radically influenced where species live and to have potentially contributed to the well-known extinctions of woolly mammoths, sabertooth cats, and enormous salamanders.

A large interdisciplinary team of scientists, including ecologists, palynologists (scientists who study pollen), paleontologists (scientists who study prehistoric life, especially fossils), climatologists, and geologists, formed a research consortium (Cooperative Holocene Mapping Project, 1988; Wright *et al.*, 1993) to study the dramatic ecological changes accompanying the transition from Ice Age to the recent interglacial period. One group of these researchers used a variety of proxy indicators to reconstruct vegetation patterns over the past 18,000 years for a significant fraction of the earth's land areas. In particular, cores of fossil pollen from dozens of sites throughout North America clearly showed how boreal tree pollen, now the dominant pollen type in the boreal zone in central Canada, was a prime pollen type during the last Ice Age (15,000–20,000 years ago) in what are now the mixed hardwood and Corn Belt regions of the United States. During the last Ice Age, most of Canada was under ice; pollen cores indicate that as the ice receded, boreal trees moved northward chasing the ice cap. One interpretation of this information was that biological communities moved intact with a changing climate. In fact, Darwin (1859) asserted as much:

> As the arctic forms moved first southward and afterward backward to the north, in unison with the changing climate, they will not have been exposed during their long migrations to any great diversity of temperature; and as they all migrated in a body together, their mutual relations will not have been much disturbed. Hence, in accordance

> with the principles inculcated in this volume, these forms will not have been liable to much modification.

If this were true, the principal ecological concern over the prospect of future climate change would be that human land-use patterns might block what had previously been the free-ranging movement of natural communities in response to climate change. The Cooperative Holocene Mapping Project, however, incorporated multiple pollen types into its analyses, including not only boreal species but also herbs and more arid (xeric) species as well as oaks and other mesic species. They discovered that during the transition from the last Ice Age to the present interglacial, nearly all species moved north, as expected. During a significant portion of the transition period, however, the distribution of pollen types provided no analog associations to today's vegetation communities (Overpeck *et al.*, 1992). That is, whereas all species moved, they did so individual by individual, not as groups. Consequently, the groupings of species during the transition period were often dissimilar to those present today (Fig. 1). The relevance of this conclusion is that in the future ecotypes will not necessarily move as a unit as climate changes (assuming there is enough time and space for such a migration). Past vegetation response to climatic change at an average rate of 1°C per millennium indicates that credible predictions of vegetation changes from comparable or even more rapid climatic changes projected from human activities cannot neglect transient (i.e., time-evolving) dynamics of the ecological system. Furthermore, caution should be exercised against relying too much on past conditions in forecasting future patterns resulting from global change forcings because the forecasted global average rate of temperature increase exceeds those rates typical of the last 120,000 years.

Future climates may not only be quite different from more recent previous climates but may also be different from those inferred from paleoclimatic data and from those to which some existing species are evolutionarily adapted. Therefore, possible future changes inferred from past changes can be taken only as a guide or a means to verify aspects of the forecasts of models of climate or ecosystem dynamics (Crowley, 1993; Schneider, 1993a). Such verification exercises may provide more credible forecasts of the effects of climatic change on animals.

III. FORCES OF CLIMATE CHANGE

The two basic categories of causes of climatic change are external and internal. However, these terms are defined relative to the focus of study; stating which components are external or internal to the climatic system depends on the time period and spatial scale being examined, as well as the phenomena being considered. External causes of climate change do not have to be physically external to the earth (such as the sun) but do occur outside of the climate system. If our focus is on atmospheric change on a 1-week time-scale (i.e., the weather), the oceans, land surfaces, biota, and human activities that produce CO_2 are all external (i.e. they are not influenced much by the atmosphere in such a short time). If our focus is on 100,000-year ice age interglacial cycles, however, the oceans, ice sheets and biota are all part of the internal climatic system and vary as an integral part of the earth's environmental systems. On this longer scale we must also include as part of our internal system the "solid" earth, which really is not solid but viscous and elastic.

Fluctuations in heat radiated by the sun—perhaps related to varying sunsports—are external to the climate system. Influences of the gravitational tugs of other planets on the earth's orbit are also external. Human-caused changes in the earth's climate could not perceptibly alter either one of these cycles.

Carbon dioxide and methane levels rise and fall with ice age cycles, which are clearly internal on a 10,000-year time-scale. However, on a 20-year scale these greenhouse gases become largely an external cause of climatic change, because small changes in climate have little feedback effect on, for example, humans burning fossil fuel.

Changes in the character of the land surface, such as those caused by human activities, are largely external. If vegetation cover changes because of climatic change, however, land surface change then becomes internal because changes in plant cover can influence the climate by changing albedo (reflectivity to sunlight), evapotranspiration, surface roughness, and relative humidity (Henderson-Sellers *et al.*, 1993).

Snow and ice are important factors in climatic change because they have higher albedo (reflectivity) than warmer surfaces and, in the instance of sea ice, can inhibit transfer of heat and moisture between air and wet surfaces. Salinity, which affects changes in sea ice and in the density of seawater (which helps control where ocean waters sink), may also be an internal cause of climatic variation. The sinking and upwelling of ocean waters are biologically significant because the upwelling waters are often nutrient-rich.

Unusual patterns of ocean surface temperature, such as the El Niño, demonstrate the importance of internally caused climatic fluctuations because the atmospheric circulation can change simultaneously with ocean sur-

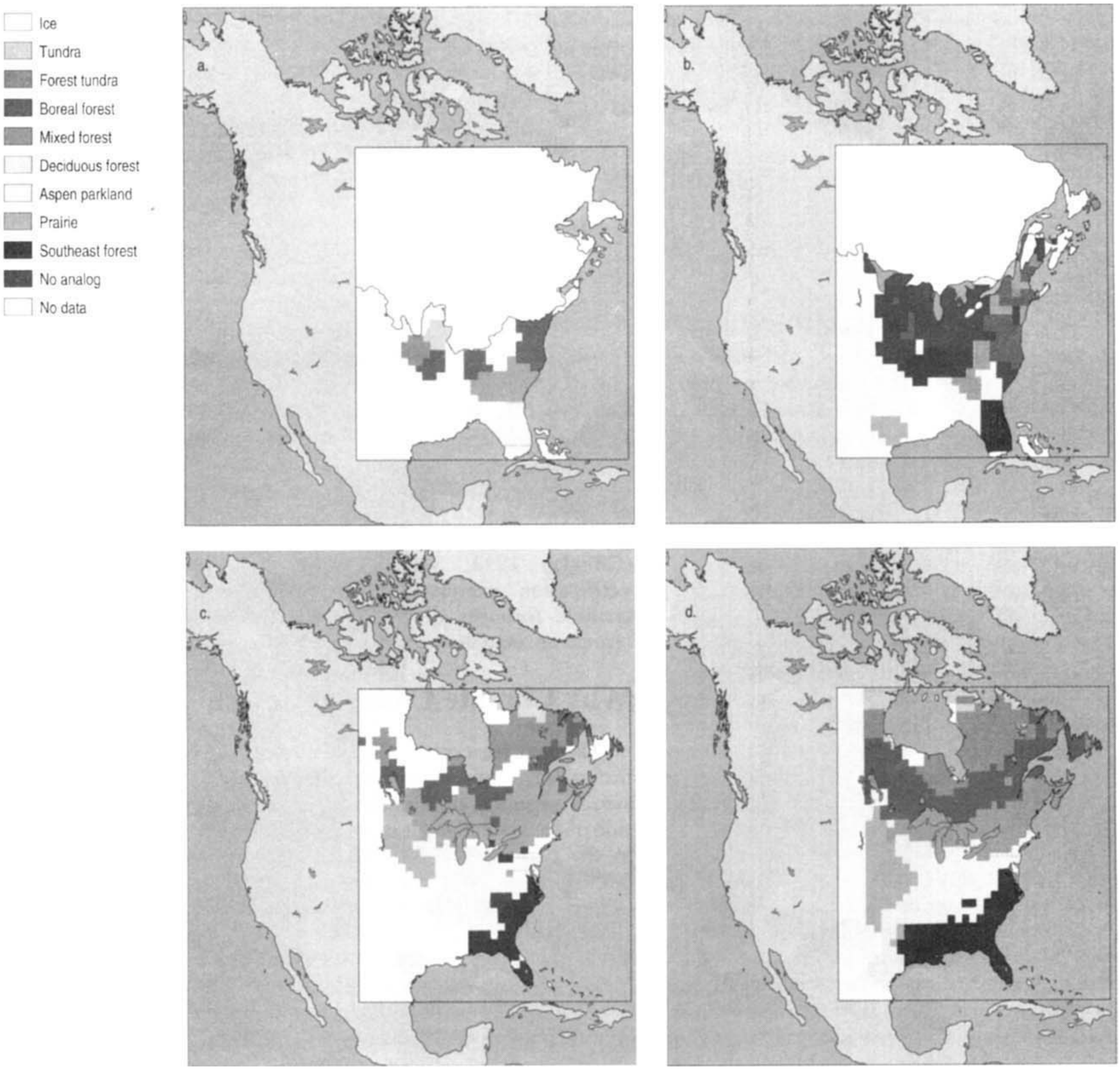


FIGURE 1 Selected paleovegetation maps reconstructed by using the method of modern analogs and more than 13,000 samples of fossils and modern pollen: (a) 18,000 BP, (b) 12,000 BP, (c) 6000 BP, and (d) modern. No vegetation is mapped in areas without fossil pollen data, and no analog refers to vegetation without any modern analog (data from Overpeck *et al.*, 1992).

face temperatures. When the atmosphere rubs on the ocean, the ocean responds by modifying its motions and temperature pattern, which forces the atmosphere to adjust, which changes the winds, which changes the way the atmosphere rubs on the ocean, and so forth (Trenberth, 1992). As a result, air and water interact internally in this coupled system like blobs of gelatin of different size and stiffness connected by elastic bands or springs, all interacting with one another while also being pushed from the outside (by solar, volcanic, or human-caused change).

IV. CLIMATE CHANGE PROJECTIONS: WHAT WILL HAPPEN?

To predict the ecologically significant ways the climate might change, one must specify what people do that modifies how energy is exchanged among the atmosphere, land surface, and space because such energy flows are the driving forces behind climate. Air pollution is an example of such a so-called societal forcing of the climate system. Estimating societal forcing involves forecasting a plausible set of human (or societal) activities affecting pollution during the next century. The next step is to estimate the response of the various components of the earth system to such societal forcings. The earth system consists of the following interacting subcomponents: atmosphere, oceans, cryosphere (snow, seasonal ice, and glaciers), and land-surface (biota and soils) systems.

Research in the field and in laboratories provides an understanding about various processes affecting the subcomponents of the earth system. This understanding can be put into mathematical expressions that, when combined, form a model of the behavior of particular components of the earth system. In practice, models of the atmosphere are connected to models of the oceans, ice, biota, and land surfaces to simulate the consequences of a scenario of societal forcing on climate and ecosystems. Controversy arises because both the societal forcing that will actually occur and the scientific knowledge of each subsystem are incomplete. Because models cannot be perfect replicas of the actual natural system, scientists must expend considerable efforts to test their models against the expanding base of field and laboratory data. This not only allows them to assess the credibility of current simulations but it also reveals improvements for the next generation of models.

A. Elements of Global Warming Forecasts

The societal driving forces behind global-warming scenarios are projections of population, consumption, land use, and technology. Typical twenty-first-century projections for human population size and affluence for less highly developed countries and more highly developed countries show drastic increases in population and wealth. When these factors are multiplied by the amount of energy used to produce a unit of economic product (the so-called energy intensity) and the amount of CO_2 emitted per unit of energy (the technology factor called carbon intensity), carbon emissions are predicted to increase severalfold during the next 100 years. It is very difficult to make such projections credibly; therefore, analysts disagree by as much as a factor of 10 about how much CO_2 will be emitted by 2100 (Johansson *et al.*, 1993; Intergovernmental Panel on Climate Change, 1996a). Specific scenarios are debatable because the amount of carbon emitted through human activities will significantly depend on what kinds of energy systems will be developed and deployed globally and on what the standards of living will be during the next several decades, not to mention population growth.

To turn estimates of CO_2 emissions into estimates of CO_2 concentrations in the atmosphere—the variable needed to calculate potential climate changes—one must estimate what fraction of CO_2 emitted will remain in the atmosphere. This airborne fraction is most simply estimated at 50% because the amount of CO_2 buildup in the atmosphere each year (about 3 billion tons of carbon as CO_2) is about half the fossil fuel-injected CO_2. The atmospheric concentration of CO_2, however, should be computed by using carbon cycles models, which account for the time-evolving amounts of carbon in vegetation, soils, and oceanic and atmospheric subcomponents (Intergovernmental Panel on Climate Change, 1996a). The estimated CO_2 concentration can then be fed into computerized climatic models to estimate its effects on climate.

Climate prediction, like most other forecasts involving complex systems, generally involves educated guesses. Those attempting to determine the future behavior of the climate system from knowledge of its past behavior and present state basically can take two approaches. One approach, the empirical–statistical, uses statistical methods such as regression equations that connect past and present observations statistically to obtain the most probable extrapolation. The second approach, usually called climate modeling, focuses on first principles, which are equations representing laws believed to describe the physical, chemical, and biological processes governing climate. Because the statistical approach depends on historical data, it is obviously limited to predicting climates that have been observed or are caused by processes appropriately represented in the past conditions. The statistical method cannot reliably answer questions such as what would happen if atmospheric CO_2 increased at rates much faster than in the known past. Thus, the more promising approach to climate prediction for conditions or forcings different from the historic or ancient past is climate modeling. A significant component of empirical–statistical information, though, is often embedded into these models. (Washington and Parkinson, 1986; Root and Schneider, 1995). This often makes modelers uncomfortable about the validity of predictions of such models on unusual or unprecedented situations unless a great deal of effort is expended to test the models against current and paleoclimatic baseline data.

Climate models vary in their spatial resolution—that is, the number of dimensions they simulate and the spatial detail they include. The simplest model calculates only the average temperature of the earth, independent of the average greenhouse properties of the atmosphere. Such a model is zero-dimensional: It reduces the real temperature distribution on the earth to a single point, a global average. In contrast, three-dimensional climate models produce the variation of temperature with latitude, longitude, and altitude. The most complex atmospheric models, the general circulation models, predict the time evolution of temperature plus humidity, wind, soil moisture, sea ice, and other variables through three dimensions in space (Washington and Parkinson, 1986).

B. Verifying Climate Forecasts

The most perplexing question about climate models is whether they can be trusted as a reliable basis for altering social policies, such as those governing CO_2 emissions or the shape and location of wildlife reserves. Even though these models are fraught with uncertainties, several methods are available for verification tests. Although no method is sufficient by itself, several methods together can provide significant, albeit circumstantial, evidence of a forecast's credibility.

The first validation testing method involves checking the model's ability to simulate the current climate. The seasonal cycle is a good test because temperature changes in a seasonal cycle are larger on a hemispheric average than the change from an ice age to an interglacial period (i.e, 15°C seasonal range in the Northern Hemisphere versus 5–7°C glacial–inter-glacial cycle). General circulation models map the seasonal cycle well. This supports the scientific consensus about the plausibility of global warming of several degrees in the twenty-first century. The seasonal test, however, does not indicate how well a model simulates slow processes such as changes in deep ocean circulation, ice cover, forests, or soil carbon storage, which may have important effects on the decade- to century-long time scales over which atmospheric CO_2 is expected to double.

A second verification technique involves isolating individual physical components of the model and testing them against actual data. A reasonable model should reproduce the flow of thermal energy among the atmosphere, the surface, and space with no more than about a 10% error. Together, these energy flows make up the well-established natural greenhouse effect on Earth and constitute a formidable and necessary test for all models. A model's performance in simulating these energy flows is an example of physical validation of model components.

A third validation method involves the model's ability to reproduce the diverse climates of the past. This method is aided by recording instrumental observations made during the past few centuries and paleo-records that serve as a proxy for climatic conditions of the ancient earth, or even include testing the models' ability to simulate climates of other planets (Kasting *et al.*, 1988). Paleoclimatic simulations of the Mesozoic (Age of the Dinosaurs), glacial–interglacial cycles, or other extreme past climates help scientists understand the coevolution of the earth's climate and living things (Schneider and Londer, 1984). As verification tests of climate models, they are also crucial to predicting future climates and changes in biological systems.

Using these techniques, much has been learned from examining the global climatic trends of the past century. The years 1997 and 1998 were the warmest on record for the lower atmosphere in the past century; at the same time, the stratosphere was at its coldest (Intergovernmental Panel on Climate Change, 1996a). These data are consistent with an enhanced greenhouse effect signal that might be anticipated from the greenhouse gas injections during the past 150 years, which saw a 25% increase in CO_2, a 150% increase in CH_4, and the introduction of human-generated heat-trapping chemicals such as chlorofluorocarbons and halons. Industrial activities since the 1950s have contributed to the increase of sulfur dioxide and other aerosol particles into the atmosphere, which reduce surface temperature by reflecting sunlight back to space. Although such cooling effects may have counteracted global warming by only several tenths of a degree, the hazes occur regionally and could be producing ecologically significant, unexpected regional changes in climate patterns (Schneider, 1994).

Although the 0.7 ± 0.2°C surface warming in the twentieth century is consistent with the human-induced greenhouse gas buildup, it is possible that the 0.7°C warming trend was wholly natural and that there was little or no contribution from the buildup of greenhouse gases (for estimates of the probability of human-induced global warming amounts, see Morgan and Keith, 1995). However, it is also not possible to rule out the counterfactual that, independent of the enhanced greenhouse effect due to human activity, there was a natural cooling fluctuation taking place during the twentieth century. If so, the world would then have warmed up much more than observed had we not had a fortuitous natural cooling trend. One could even speculate that the dramatic temperatures since the 1970s with global high temperature records reflect the termination of a natural cooling trend combined with the rapid establishment of the expected enhanced greenhouse effect.

Studies (Santer *et al.*, 1996) suggest that when aerosols and greenhouse gas forcings are combined, climate models more closely match 30 years of observations. Nevertheless, there is wide variability in predictions for a doubling of CO_2: Temperature changes as low as a 0.5°C warming to as high as a 5.0°C warming (Wigley and Raper, 1991) are all consistent with current observations. Several reasons exist for such a wide range of uncertainty: difficulty in knowing how to model delays in global warming because of the large heat capacity of the oceans; not knowing what other global change forcings may have opposed warming, for example, sulfate aerosols from burning high-sulfur coal and oil or undetectable changes in the sun's light output before 1980; and large, unknown, internal natural climatic fluctuations. As mentioned previously, though, the ecologically important forecasts of time-evolving regional climatic changes are much less credible and require that ecologists use many alternative scenarios of possible climatic changes.

In summary, no clear physical objection or direct empirical evidence have contradicted the consensus of scientists (Intergovernmental Panel on Climate Change, 1990, 1996a) that the world is warming, nor has evidence emerged to contradict the substantial probability that temperatures will rise because of increases in greenhouse gases (Morgan and Keith, 1995). The evidence for current global warming forecasts is circumstantial, but it is sufficient enough that many researchers believe that recently observed climatic variations and human activities are probably related (Karl *et al.*, 1995). The Intergovernmental Panel on Climate Change (1996a) carefully weighed the uncertainties and concluded that "the balance of evidence suggests that there is a discernible human influence on global climate."

V. RELEVANCE OF CLIMATE MODELING TO REGIONAL CLIMATE CHANGE AND ECOSYSTEM STUDIES

Scientists who estimate the future climatic changes that are relevant to ecosystems have focused on the general circulation models that attempt to represent mathematically the complex physical and chemical interactions among the atmosphere, oceans, ice, biota, and land. As these models have evolved, increasingly more information has become available, and more comprehensive simulations have been performed. Nevertheless, the complexities of the real climate system still vastly exceed the general circulation models and the capabilities of even the most advanced computers (Intergovernmental Panel on Climate Change, 1990, 1996a). Simulating 1 year of weather in 30-min time steps with the crude resolution of 40 latitudinal lines × 48 longitudinal lines and 10 vertical layers—nearly 20,000 grid cells throughout the world—takes several hours on a supercomputer. This level of resolution, however, cannot resolve the Sierra Nevada of California and the Rocky Mountains as separate mountain chains. Refining the resolution to 50-km grid squares would so dramatically increase the number of computations that it would take approximately months of computer time to simulate weather statistics for 1 year.

Even the highest-resolution, three-dimensional general circulation model will not have a grid with nodes much less than 100 km apart within the foreseeable future; individual clouds and most ecological research (to say nothing of cloud droplets) occur on scales far smaller. Therefore, general circulation models will not be able to resolve the local or regional details of weather affecting most local biological communities or the importance of regional effects of hills, coastlines, lakes, vegetation boundaries, and heterogeneous soil (Root and Schneider, 1993). Nonetheless, it is important to have climatic forecasts and ecological response analyses on the same physical scales.

What is most needed to evaluate potential biological effects of temperature change is a regional projection of climatic changes that can be applied to ecosystems at a regional or local scale. Analyses of large, prehistoric climatic changes (Barron and Hecht, 1985; Budyko *et al.*, 1987; Schneider, 1987; Cooperative Holocene Mapping Project, 1988) and historical weather analogs (Pittock and Salinger, 1982; Jager and Kellogg, 1983; Lough *et al.*, 1983; Shabalova and Können, 1995) provide some insights into such changes. Historical weather analogs, however, since they are empirically and statistically based, rely on climatic cause-and-effect processes that probably differ from those that will be driven by future greenhouse gas radiative effects (Schneider, 1984; Mearns *et al.*, 1990; Crowley, 1993). Consequently, ecologists turn to climatic models to produce forecasts of regional climatic changes for the decades ahead. How credible are such forecasts?

A. Regional Changes

Although the consensus among researchers about the plausibility of significant human-induced global climatic change is growing, no assessment (Intergovernmental Panel on Climate Change, 1996a) has suggested the existence of a strong consensus about how that global climatic change might be distributed regionally. For example, the world is not actually undergoing a dramatic and instantaneous doubling of CO_2, which is the hypothesis used in most standard computer model experiments applied to ecological assessments. Instead, the world is undergoing a steady increase in greenhouse gas forcing. Because that increase is heating the earth in a reasonably uniform way, one might expect a uniform global response, although this is far from likely. For example, the centers of continents have relatively low heat-retaining capacity, and the temperatures there would move relatively rapidly toward whatever their new equilibrium climate would be compared with the centers of oceans, which have high heat-retaining capacity. Tropical oceans, though, have a thin (about 50 m) mixed layer that interacts primarily with the atmosphere. It takes about 10 years for that mixed layer to substantially change its temperature, which is still much slower than the response time of the middle of the continents but is much faster than that of the oceans closer to the poles. At high latitudes, in places such as the Weddell or Norwegian seas, waters can mix down to the bottom of the ocean, thereby continuously bringing up cold water and creating a deep-water column for which a century or more is required to substantially change its temperature.

During the transient phase of climate change over the next century, therefore, one would expect the middle of continents, the middle of oceans, and the polar and subpolar oceans all to change toward their new equilibrium temperatures at different rates. Thus, the temperature differences from land to sea and equator to pole will evolve over time, which in turn implies that the transient character of regional climatic changes could be very different from the expected long-term equilibrium (Schneider and Thompson, 1981; Stouffer *et al.*, 1989; Washington and Meehl, 1989). This does not imply that transient regional changes are inherently unpredictable, only that currently they are very difficult to predict credibly.

Even more uncertain than regional averages, but perhaps more important to long-term ecosystem responses, are estimates of climatic variability during the transition to a new equilibrium, particularly at the regional scale. These include estimates of events such as the frequency and magnitude of severe storms, enhanced heat waves, temperature extremes, sea-level rises (Titus and Narayanan, 1995), and reduced frost probabilities (Mearns, *et al.*, 1984, 1990; Parry and Carter, 1985; Wigley, 1985; Rind *et al.*, 1989). For example, there is a physical principle that evaporation increases dramatically as surface-water temperature increases. Because hurricanes are powered by evaporation and condensation of water, if all other factors are unchanged, the intensity of hurricanes and the length of the hurricane season could increase with warming of the oceans (Emanuel, 1987; Knutson, 1998). Such changes would significantly affect susceptible terrestrial and marine ecosystems (Doyle, 1981; O'Brien *et al.*, 1992).

B. Downscaling Climate Predictions to Regional Effects

1. Empirical Mapping Techniques

Techniques exist that can translate the output of climate models so that it is closer to that of most ecological scales. One method that uses actual climatic data at both large and small scales can help provide maps that may allow small-scale analysis of large-scale climate change scenarios. For example, the Sierra Nevada of California or the Cascades in the northwestern United States are north–south mountain chains whose east–west dimensions are smaller than the grid size of a typical general circulation model. In the actual climate system, onshore winds on the Pacific coast would produce cool upslope and rainy conditions on the western slope and a high probability of warmer and drier conditions associated with that flow pattern on the downslope or eastern slope.

One regional map has been generated for Oregon (Gates, 1985) in which a high-resolution network of meteorological stations was used to plot temperature and precipitation isopleths based on observed climatic fluctuations at large (e.g., state-sized) scales. These maps show that the dominant mode of variation for this area is warm and dry on one side of the mountains and cold and wet on the other side. Although this empirical mapping technique seems appropriate for translating low-resolution, grid-scale climate model forecasts to local applications, a strong caveat must be offered. That is, the processes in the climate system that give rise to internal variability or natural fluctuations are not necessarily the same processes that would give rise to local deviations from large-scale patterns if the climate change were driven by external forces rather than an internal variation of the system. For example, the Oregon maps would indicate that if the grid box average temperature were warmer on the eastern slope, then it should be cooler and wetter on the western slope. This condition is the most probable regional situation for today's naturally fluctuating climate. However, if 50 years from now the warming on the eastern slope were, for example, a result of doubled atmosphere CO_2 causing an enhanced downward infrared radiative heating, then both eastern and western slopes would probably experience warming. Although the degree of warming and associated precipitation changes would not necessarily be uniform, an entirely different climatic change pattern would probably occur as opposed to that obtained from the empirical mapping technique if one used the naturally varying weather conditions existing today rather than the anthropogenically forced conditions of the twenty-first century (Schneider, 1993b).

Therefore, techniques to shrink climate forecasts that use current distributions of environmental variables at local scales and correlate them with current large-scale regional patterns will not necessarily provide a good guideline about how large-scale patterns would be distributed regionally. The reason is that the causes of the future change may be physically or biologically different from the causes of the historical fluctuations that led to the empirical maps in the first place. This caveat is so important that it requires scientists to use extreme caution before adopting such empirical techniques for global change applications.

2. Regional-Scale Models with General Circulation Model Inputs

Other techniques can still translate large-scale patterns to smaller scales, but these techniques are based on known processes rather than empirical maps for today's conditions. One such technique is to drive a high-resolution, process-based model for a limited region with the large-scale patterns produced by a general circulation model. In essence, this approach uses a mesoscale model (i.e., 10–50-km^2 grid cells) based on physical laws to solve the problem of translating general circulation model grid-scale averages into a finer scale mesh much closer to the dimensions of most ecological applications. Of course, even this mesoscale grid will still be too coarse to assess many impacts, necessitating further downscaling techniques. Neither are the problems of general circulation models entirely eliminated by mesoscale grids because they are bigger than individual clouds or trees. However, such methods do bring climate model scales and ecological-response scales much closer.

VI. EXAMPLES OF ECOLOGICAL RESPONSES TO CLIMATE CHANGES

Bringing climatic forecasts to ecological applications at local and regional scales is one way to bridge the scale gap across ecological and climatological studies. Ecologists, however, have also analyzed data and constructed models that apply over large scales, including the size of climatic model grids. A long tradition in ecology has associated the occurrence of vegetation types or the range limits of different species with physical factors such as temperature, soil moisture, land–sea boundaries, or elevation. Biogeography is the field that deals with such associations, and its results have been applied to estimate the large-scale ecological response to climate change.

A. Predicting Vegetation Responses to Climate Change

The Holdridge (1967) life zone classification assigns biomes (e.g., tundra, grassland, desert, or tropical moist forest) according to two measurable variables—temperature and precipitation. Other more complicated large-scale formulas have been developed to predict vegetation patterns from a combination of large-scale predictors (e.g., temperature, soil moisture, or solar radiation); vegetation modeled includes individual species (Davis and Zabinski, 1992), limited groups of vegetation types (Box, 1981), or biomes (Prentice, 1992; Melillo *et al.*, 1993; Neilson, 1993). These kinds of models predict vegetation patterns that represent the gross features of actual vegetation patterns, which is an incentive to use them to predict vegetation change with changing climate, but they have some serious drawbacks as well. That is, they are typically static, not time-evolving dynamic simulations, and thus cannot capture the transient sequence of changes that would take place in reality. In addition, such static biome models occasionally make "commission errors"; that is, they predict vegetation types to occur in certain zones where climate would indeed permit such vegetation, but other factors such as soils, topography, or disturbances such as fire actually preclude it. Furthermore, local patterns may influence vegetation dynamics at scales not captured in some simulations, and seed germination and dispersal mechanisms are also either not explicitly simulated or simulated only crudely with such models. It is remarkable that they are still able to produce generalized maps of vegetation types which do indeed resemble current or even paleoclimatic patterns in a broad sense, but their details do not provide confident projections for future vegetation states. Fortunately, progress is being made to include some of the deficiencies mentioned previously, and so-called dynamical global vegetation models are being developed to treat the transient nature of vegetation change that would likely accompany climatic change.

B. Predicting Animal Responses to Climate Change

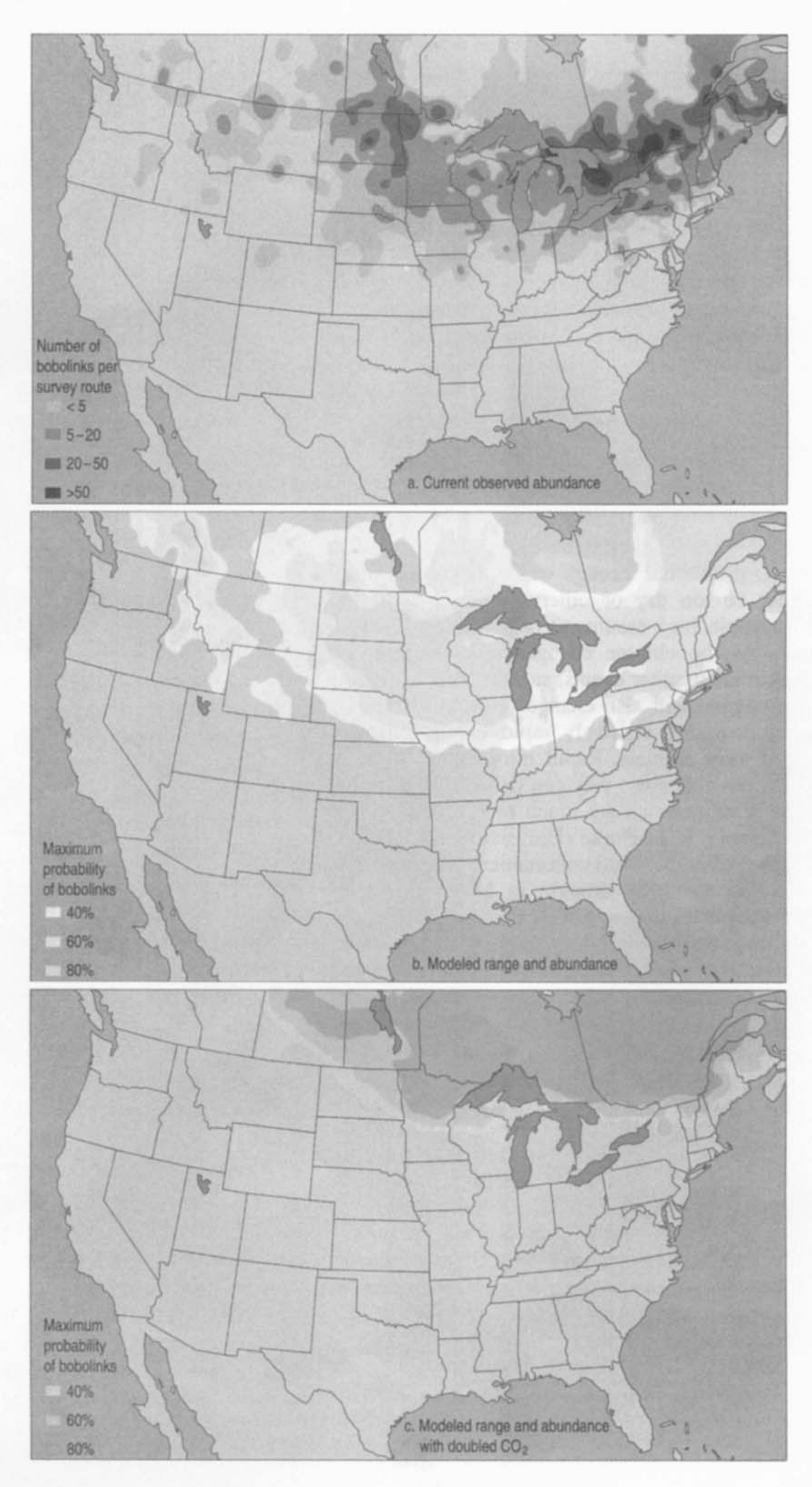
Scientists of the U.S. Geological Survey, in cooperation with Canadian scientists, conduct the annual North American Breeding Bird Survey, which provides distribution and abundance information for birds throughout the United States and Canada. From these data, collected by volunteers under strict guidance from the U.S. Geological Survey, shifts in bird ranges and abundances can be examined. Because these censuses were begun in the 1960s, these data can provide a wealth of baseline information. Price (1995) used these data to examine the birds that breed in the Great Plains. By using the present-day ranges and abundances for each of the species (Fig. 2a), Price derived large-scale, empirical–statistical models based on various climate variables (e.g., maximum temperature in the hottest month and total precipitation in the wettest month) that provided estimates of the current bird ranges and abundances (Fig. 2b). Then, by using a general circulation model to forecast how doubling of CO_2 would affect the climate variables in the models, he applied the statistical models to predict the possible shape and location of the birds' ranges and abundances (Fig. 2c).

Significant changes were found for nearly all birds examined. The ranges of most species moved north, up mountain slopes, or both. The empirical models assume that these species are capable of moving into these more northerly areas, provided habitat is available and no major barriers exist. Such shifting of ranges and abundances could cause local extinctions in the more southern portions of the birds' ranges, and, if movement to the north is impossible, extinctions of entire species could occur. We must bear in mind, however, that this empirical–statistical technique, which associates large-scale patterns of bird ranges with large-scale patterns of climate, does not explicitly represent the physical and biological mechanisms that could lead to changes in birds' ranges. Therefore, the detailed maps should be viewed only as illustrative of the potential for very significant shifts with different possible doubled CO_2 climate change scenarios. More refined techniques that also attempt to include actual mechanisms for ecological changes are discussed later.

Reptiles and amphibians, which together are called herptiles, are different from birds in many ways that are important to our discussion. First, because herptiles are ectotherms—meaning that their body temperatures adjust to the ambient temperature and radiation of the environment—they must avoid environments in which temperatures are too cold or too hot. Second, amphibians must live near water not only because the reproductive part of their life cycle is dependent on water but also because they must keep their skin moist because they breathe through their skin as well as their lungs. Third, herptiles are not able to disperse as easily as birds because they must crawl rather than fly, and the habitat through which they crawl must not be too dry or otherwise impassable (e.g., high mountains or superhighways).

As the climate changes, the character of extreme weather events, such as cold snaps and droughts, will also change (Karl *et al.*, 1995), necessitating relatively rapid habitat changes for most animals. Rapid movements by birds are possible since they can fly, but for herptiles such movements are much more difficult. For example, R. L. Burke (personal communication) noted that during the 1988 drought in Michigan, many more turtles than usual were found dead on the roads. He assumed they were trying to move form their usual water holes to others that had not yet dried up or that were cooler. For such species, moving across roads usually means high mortality. In the long term, most birds can readily colonize new habitat as cliamtic regimes shift, but herptile dispersal (colonization) rates are slow. Indeed, some reptile and amphibian species may still be expanding their ranges north even now, thousands of years after the last glacial retreat.

R. L. Burke and T. Root (personal communication) have begun analyzing North American herptile ranges in an attempt to determine which, if any, are associated with climatic factors such as temperature, vegetation-greening duration, and solar radiation. Their preliminary evidence indicates that northern boundaries of some species ranges are associated with these factors, implying that climatic change could have a dramatic impact on the occurrence of herptile species. It could also alter the population genetics within species since there can be genetic differences among populations with respect to climate tolerance. Furthermore, several reptile and most North American turtle species could exhibit vulnerability to climatic change because the temperature experienced as they develop inside the egg determines their sex. Such temperature-dependent sex determination makes these animals uniquely sensitive to temperature change, meaning that climatic change could potentially cause dramatic range contractions. Many more extinctions are possible in herptiles than in birds because the forecasted human-induced climatic changes could occur rapidly when compared with the rate of natural climatic changes, and because the dispersal ability of most herptiles is very slow, even without considering the additional difficulties associated with human land-use changes disturbing their migration paths.


Number of bobolinks per survey route
<5
5–20
20–50
>50
a. Current observed abundance
Maximum probability of bobolinks
40%
60%
80%
b. Modeled range and abundance
Maximum probability of bobolinks
40%
60%
80%
c. Modeled range and abundance with doubled CO₂

In general, animals most likely to be affected earliest by climatic change are those for which populations are fairly small and limited to isolated habitat islands. As a result of human-generated landscape changes, many reptiles now fall into this category, as do many other animals. Indeed, temperature-dependent sex-determined species are especially likely to suffer from extreme sex ratio biases, and therefore their sensitivity to rapid climate change appears potentially more severe than that of most other animals.

There are estimates that many small mammals living near isolated mountaintops (which are essentially habitat islands) in the Great Basin would become extinct given typical global change scenarios (MacDonald and Brown, 1992). Recent studies of small mammals in Yellowstone National Park show that statistically significant changes in both abundances and physical sizes of some species occurred with historical climate variations (which were much smaller than most projected climate changes for the next century), but there appear to have been no simultaneous genetic changes (Hadley *et al.*, 1997). Therefore, it is likely that climate change in the twenty-first century could cause substantial alteration to biotic communities, even in protected habitats such as Yellowstone National Park. In addition, the biomass of macrozooplankton in waters off southern California has decreased dramatically as surface waters warmed (Roemmich and McGowan, 1995). Similarly, a study suggests that statistically the range of the Edith's checkerspot butterfly in western North America has shifted northward and upward in association with long-term regional warming trends (Parmesan, 1996).

C. Top-Down Approaches

The biogeographic approach previously summarized is an example of a top-down technique (e.g., Holdridge life zone classification), in which data on abundances or range limits of vegetation types or biomes are overlain on data of large-scale environmental factors such as temperature or precipitation. When associations among large-scale biological and climatic patterns are revealed, biogeographic rules expressing these correlations graphically or mathematically can be used to forecast changes in vegetation driven by given climate changes.

D. Bottom-Up Approaches

Another traditional analysis and forecasting technique is often referred to as bottom-up. Small-scale ecological studies have been undertaken at the scale of a plant or even a single leaf (Idso and Kimball, 1993) to understand how, for example, increased atmospheric CO_2 concentrations might directly enhance photosynthesis, net primary production, or water-use efficiency. Most such studies indicate increases in all these factors—increases that some researchers have extrapolated to ecosystems (Idso and Brazel, 1984; Ellsaesser, 1990).

However, at the scale of a forest, the relative humidity within the canopy, which significantly influences the evapotranspiration rate, is regulated by the forest. In other words, if an increase in water-use efficiency decreased the transpiration from each tree, the aggregate forest effect would be to lower relative humidity. This, in turn, would increase transpiration, thereby offsetting some of the direct CO_2/water-use efficiency improvements observed experimentally at the scale of a single leaf or plant. Regardless of the extent to which this forest-scale feedback effect will offset inferences made from bottom-up studies of isolated plants, the following general conclusion emerges: The bottom-up methods may be appropriate for some processes at some scales in environmental science, but they cannot be considered credible without some sort of validation testing at the scale of the system under study.

E. Combined Top-Down and Bottom-Up Approaches

To help resolve the deficiencies of the top-down biome forest models mentioned previously, more process-based, bottom-up approaches such as forest gap models have been developed (Botkin *et al.*, 1972; Pastor and Post, 1988; Smith *et al.*, 1992). These models include individuals species and can calculate vegetation dynamics driven by time-changing climatic change scenarios.

FIGURE 2 (a) Map of current range and abundance of the bobolink as determined from actual observations during the U.S. Geological Survey Breeding Bird Survey and (b) map of current range and abundance of the bobolink as estimated from the empirical–statistical model. The high correspondence in patterns between the maps in a and b suggests that this model reliably captures many of the features of the actual observed range and abundance of this species as depicted in the map in a. (c) Map of the forecasted range and abundance of the bobolink for climate change response of a model with doubled CO_2. This map illustrates the potential for very significant shifts that doubled CO_2 would cause.

However, the actual growth rate calculated in the model for each species has usually been determined by multiplying the ideal growth rate curve by a series of growth-modifying functions that attempt to account for the limiting effects of nutrient availability, temperature stress, and so forth. These growth-modifying functions for temperature are usually determined empirically at a large scale by fitting an upside-down U-shaped curve, whose maximum is at the temperature midway between the average temperature of the species' northern range limit and the average temperature of its southern range limit. Growing degree-days (related to temperature but not temperature per se) are used in this scenario.

In essence, this technique combines large-scale, top-down empirical pattern correlations into an otherwise mechanistic bottom-up modeling approach. Although this combined technique refines both approaches, it has been criticized because such large-scale, top-down inclusions are not based on the physiology of individual species and lead to confusion about the fundamental and realized niches (Pacala and Hurtt, 1993). (The fundamental niche is the ecological space in which a given species could theoretically survive—for example, if its competitors were absent—and the realized niche is where it actually exists). The question is then, what limits the realized niche, particularly at the southern boundary? Furthermore, more refined models should include factors such as seed dispersal so that plant recruitment is related to the preexisting population and is not simply the result of a random number generator in the computer code.

F. Studies of More Refined Approaches

As noted previously, problems with the singular use of either top-down or bottom-up methods have led to well-known criticisms. For bottom-up models, the primary problem is that some of the most conspicuous processes observable at the smaller scales may not be the dominant processes that generate large-scale patterns.

Top-down approaches suffer because of the possibility that the discovered associations at large scales are statistical artifacts that do not, even implicitly, reflect the causal mechanisms needed for reliable forecasting. As Jarvis (1993, p. 121) stated, "A major disadvantage of a top-down model is that predictions cannot be made safely outside the range of the variables encountered in the derivation of the lumped parameter function."

A search of the literature (Wright *et al.*, 1993; Root, 1994; Harte *et al.*, 1995) provides examples of a refined approach to analyzing across large and small scales, which Root and Schneider (1995) labeled strategic cyclical scaling. This method builds on the combined techniques in which top-down and bottom-up approaches are applied cyclically in a strategic design that addresses a practical problem: in our context, the ecological consequences of global climatic change. Large-scale associations are used to focus small-scale investigations; this helps ensure that tested causal mechanisms are generating the large-scale relations. Such mechanism become the laws that allow more credible forecasts of the consequences of global change disturbances. Levin (1993) observed.

> Although it is well understood that correlations are no substitute for mechanistic understanding of relationships, correlations can play an invaluable role in suggesting candidate mechanisms for (small-scale) investigation. (p. 14)

Strategic cyclical scaling, however, is intended not only as a two-step process but also as a continuous cycling process between large- and small-scale studies, with each successive investigation building on previous insights from all scales. This approach is designed to enhance the credibility of the overall assessment process (Vitousek, 1993; Harte and Shaw, 1995), which is why strategic is the first word in strategic cyclical scaling.

If the rate at which humans are injecting greenhouse gases into the atmosphere is not greatly decreased, there is a significant chance that the earth's climate will warm by several degrees Celsius by 2050 (Titus and Narayanan, 1995). With this in mind, Root (1988) examined the biogeographic patterns of all wintering North American birds. She chose this group of species because birds are important parts of ecosystems and because of the availability of the necessary data. The National Audubon Society and the U. S. Geological Survey have volunteer forces amassed to aid in the collection of Christmas Bird Count data and Breeding Bird Survey data, respectively. By using Christmas Bird Count data, Root determined that for a large proportion of species average distribution and abundance patterns are associated with various environmental factors (e.g., the northern range limits of some species apparently may be limited by average minimum January temperature; Root, 1988, 1989; Repasky, 1991).

The following is the scaling question: What mechanisms (such as competition or thermal stress) at small scales may have given rise to the large-scale associations? Root first tested the hypothesis that local physiological constraints may be causing most of the particular large-scale, temperature-range boundary associations. She used published small-scale studies on the wintering physiology of key species to determine that about half of the songbird species wintering in North America extend their ranges no farther north than the regions where, to avoid hypothermia during winter nights, they need not increase their metabolic rates more than approximately 2.5 times their basal metabolic rate (Root, 1988). Root embarked on a larger, regional study to determine whether the longer nights—hence, fewer hours of daylight available for foraging—or the colder temperatures in the more northerly locations are relatively more important. Preliminary results indicate that changing temperatures are more likely than day length to explain this effect (Root, unpublished data): Thus, global temperature changes would probably cause a rapid range and abundance shift, at least by selected bird species. Indeed, Root found significant year-to-year shifts in ranges and abundances; these shifts are apparently associated with year-to-year changes in winter temperatures. No claim is made at this point in the research for the generally of the preliminary results indicating strong and quantitative links between bird disturbances and climate. This example does permit, however, a clear demonstration of refined methods for cycling across scales to estimate ecological responses to climatic change.

VII. THREE-WAY LINKAGES AND COMMUNITY ECOLOGY

The anticipated changes in plant ranges will probably have dramatic effects on animals, both on the large biogeographic scale and on the local regional scale. The ranges of many animals are strongly linked to vegetation. For example, red-cockaded woodpeckers are endemic to mature longleaf pine and pine–oak forests (Mengel and Jackson, 1977), and the winter range of Sprague's pipit is coincident with bluestem, a grass (Root, 1988). Consequently, the ranges of various animals that rely on specific vegetation will change as the ranges of these plants shift, assuming that some other factor is not limiting them. If the climate changes more rapidly than the dispersal rates of the plants, it will result in extensive plant die-offs in the south or down slope before individuals can disperse and become established in the north or upslope. Thus, the ranges of animals relying on these plants could become compressed, and in some instances both the plants and the animals could become extinct. For instance, the red-cockaded woodpecker needs mature, living trees for nesting sites (Jackson, 1974), and if increasing temperature causes most large trees to die before the newly established dispersing trees grow large enough, then this woodpecker, federally listed as endangered, could easily become extinct.

Many animal species have ranges that are not directly limited by vegetation but are instead restricted by temperature. This is true for most ectotherms (insects and related arthropods, amphibians, and reptiles) as well as some endotherms (mammals and birds). For example, the eastern phoebe, a North American songbird, winters in the United States in areas with average minimum temperatures warmer than 4°C (Root, 1988). As the earth warms, those species directly limited by temperature will be able to expand northward as rapidly as their dispersal mechanisms will allow, again assuming other factors are not limiting them. The animals limited by vegetation will be able to expand their ranges only as rapidly as the vegetation changes. Consequently, the potential for significant disruption among communities is high. For instance, some animals may no longer be able to coexist because an invading species disrupts the balance between competing species or between predator and prey. Therefore, to understand the ecological consequences of global climatic change on animals, the three-way linkages among animals, plants, and climate must be understood. It is critical to realize that this is not simply a one-way process whereby climate influences biota but rather a three-way process because animals and plants affect each other and are affected by climate. At the same time, altered surface vegetation can affect climate because mid-continental summer precipitation is significantly influenced by water vapor from evapotranspiration (Ye, 1989; Salati and Nobre, 1991).

VIII. CLIMATE FORECASTS, ECOSYSTEM RESPONSES, AND SYNERGISTIC EFFECTS

A. Improve Regional Analysis, Study Transients, and Include Many Variables

The most reliable projections from climatic models are for global-scale temperature changes. Ecological impact assessments, however, need time-evolving (transient) scenarios of regoinal to local-scale climate changes: included are changes in precipitation; severe storm intensity, frequency, and duration; drought frequency, intensity, and duration; soil moisture; frost-free days; intense heat waves; ocean currents; upwelling zones; near-ground ozone; forest canopy humidity; and ultraviolet radiation and total solar radiation reaching the surface, where photosynthesis is important. Data gathered at many scales and by coordinated volunteer and professional sources are needed for archives of these regional and local variables, which in turn can be used to develop and test models or other techniques for climatic forecasting.

B. Abrupt Climatic Changes

We have argued that sustained globally averaged rates of Earth and ocean surface temperature changes from the past Ice Age to the present were about 1°C per 1000 years. Alarmingly, this is a factor of approximately 10 slower than the expected changes of several degrees Celsius per 100 years typically projected for the twenty-first century due to human effects. We emphasize the words sustained globally averaged because comparably rapid regional variations have occurred. For example, about 13,000 years ago, after warm-weather fauna had returned to northern Europe and the North Atlantic, there was a dramatic return to ice age-like conditions in less than 100 years. This Younger Dryas miniglacial lasted about 1000 years before the stable recent period was established (Berger and Labeyrie, 1987). The Younger Dryas was also accompanied by dramatic disturbances to plants and animals in the North Atlantic and Europe (Coope, 1977; Ruddiman and McIntyre, 1981). During the same period, dramatic shifts can be found outside of the North Atlantic region (e.g., Sevringhaus, 1998), but no significant climate change is evident in Antarctic ice cores. However, studies of fossils in the North Atlantic show that the warm gulf stream current deviated many degrees of latitude to the south and that the overall structure of deep-ocean circulation may have returned to near ice age form in only decades—a weakening of the vertical circulation known as the conveyor-belt current (Broecker *et al.*, 1985).

Plausible speculations about the cause of the Younger Dryas center on the injection of fresh meltwater into the North Atlantic, presumably associated with the breakdown of the North American ice sheet (Boyle and Weaver, 1994; Paillard and Labeyrie, 1994). Could such a rapid change to the conveyor-belt current be induced today by pushing the present climatic system with human disturbances such as greenhouse gases or sulfur oxides? The potential for this is speculative, of course, but its possibility has concerned many scientists (Broecker, 1994, 1998; Rahmstorf, 1999). The prospect of climatic surprises in general is chilling enough to lend considerable urgency to the need to speed up the rate of our understanding, slow down the rates at which we are forcing nature to change, or both.

C. Adaptability

Our current inability to credibly predict time-evolving regional climatic changes has many implications, one of which concerns the adaptability of agricultural ecosystems. That is, any experience farmers might have with anomalous weather in, for example, the 2020s may not help them adapt to the evolving climate change in the 2030s because a transient climate change could differ dramatically over time. This would inhibit learning by doing, creating a potential lack of adaptability associated with the difficulty of reliably predicting regional climatic consequences (Schneider *et al.*, 2000). Such rapid climate changes would be especially difficult for natural ecosystems to adapt to because habitats do not have the luxury of "choosing" to plant new seeds or change irrigation systems, soil tillage practices, or other agricultural practices.

D. Ecological Applications-Driven Climatic Research

Regional projections of climatic change arising from a variety of greenhouse gas and sulfur oxide emissions scenarios are essential for ecological applications. Such studies must stress the climatic variables most likely to have significant effects on biological resources. For example, extreme variability measures such as high temperature and low relative humidity are important for evaluating the risk of forest fires (Torn and Fried, 1992). Identifying such variables of ecological importance and communicating this information to climate scientists require close interdisciplinary, multi-institutional, and cross-scale research efforts to ensure that combinations of variables relevant to ecological applications receive research priority by climatologists. A focus of climate research toward changing climatic variability (Mearns *et al.*, 1984, 1990; Rind *et al.*, 1989) might be more useful for ecological impact assessments than the current focus among climatic modelers on climatic means.

E. Interactive, Multiscale, Ecological Studies Needed

Most ecological studies project the response of one species at small scales or shifts in biomes at large scales to an equilibrium, CO_2-doubled climate model (e.g., the

Vegetation/Ecosystem Modeling and Analysis Project, 1995). What is needed for more realistic and useful ecological impact assessments is a multi-scale, multi-species, multi-taxa analysis driven by regionally specific, transient climatic change forecasts. The construction of ecological forecasts models first requires large-scale data sets gathered locally by professional (e.g., U.S. Geological Survey landcover data sets) and volunteer (e.g., National Audubon Society Christmas Bird Count) workers. Without such data sets, virtually no credible progress is possible in determining large-scale patterns of associations among ecological and climatic variables. Small-scale studies informed by large-scale patterns are then needed to refine causal mechanisms underlying such large-scale associations, thereby testing the formulas used to make projections of various species or biome responses to hypothesized global changes. For example, Pacala and Hurtt (1993) suggested small- to medium-scale experiments to improve forest gap models. Their criticisms suggest that largely first principles, bottom-up models may still be unrealistic if some top-down parameters (i.e., growth-modifying functions in the instance of gap models) are not appropriately derived from data at the scale at which the models are being applied (Root and Schneider, 1995).

One obvious truism emerges: Credible modeling required for forecasting across many scales and for complex interacting systems is a formidable task requiring repeated testing of many approaches. Nevertheless, tractable improvements in refining combined top-down and bottom-up techniques can be made. It will, however, take more than one cycle of interactions to reliably address the cross-scale and multi-component problems of ecological assessment—what we (Root and Schneider, 1995) have labeled strategic cyclical scaling.

See Also the Following Articles

CARBON CYCLE • CLIMATE, EFFECTS OF • DIVERSITY, COMMUNITY/REGIONAL LEVEL • ENERGY USE, HUMAN • GEOLOGIC TIME, HISTORY OF BIODIVERSITY IN • GREENHOUSE EFFECT

Bibliography

Davis, M. B., and Zabinski, C. (1992). Changes in geographical range resulting from greenhouse warming effects on biodiversity in forests. In *Global Warming and Biological Diversity* (R. L. Peters and T. E. Lovejoy, Eds.), pp. 297–308. Yale Univ. Press, New Haven, CT.

Graham, R. W. (1992). Late Pleistocene faunal changes as a guide to understanding effects of greenhouse warming on the mammalian fauna of North America. In *Global Warming and Biological Diversity* (R. L. Peters and T. E. Lovejoy, Eds.), pp. 76–87. Yale Univ. Press, New Haven, CT.

Holdridge, L. R. (1967). *Life Zone Ecology*. Tropical Science Center, San Jose, Costa Rica.

Intergovernmental Panel on Climate Change. (1996a). *Climate Change 1995—The Science of Climate Change. The Second Assessment Report of the IPCC: Contribution of Working Group I* (J. T. Houghton, L. G. Meira Filho, B. A. Callander, N. Harris, A Kattenberg and K. Maskell, Eds.). Cambridge Univ. Press, Cambridge, UK.

Intergovernmental Panel on Climate Change. (1996b). *Climate Change 1995—Impacts, Adaptations and Mitigation of Climate Change: Scientific Technical Analysis. The Second Assessment Report of the IPCC: Contribution of Working Group II* (R. T. Watson, M. C. Zinyowera and R. H. Moss, Eds.). Cambridge Univ. Press, Cambridge, UK.

Intergovernmental Panel on Climate Change. (1996c). *Climate Change 1995—Economic and Social Dimensions of Climate Change. The Second Assessment Report of the IPCC: Contribution of Working Group III* (J. Bruce, H. Lee and E. Haites, editors. Cambridge Univ. Press, Cambridge, UK.

Levin, S. A. (1993). Concepts of scale at the local level. In *Scaling Physiological Processes: Leaf to Globe* (J. R. Ehleringer and C. B. Field, Eds.), pp. 7–19. Academic Press, New York.

Morgan, M. G., and Keith, D. W. (1995). Subjective judgements by climate experts. *Environ. Sci. Technol.* 29, 468A–477A.

Overpeck, J. T., Webb, R. S., and Webb, T. III (1992). Mapping eastern North American vegetation change over the past 18,000 years: No analogs and the future. *Geology* 20, 1071–1074.

Pacala, S. W., and Hurtt, G. C. (1993). Terrestrial vegetation and climate change: Integrating models and experiments. In *Biotic Interactions and Global Change* (P. Kareiva, J. Kingsolver and R. Huey, Eds.), pp. 57–74 Sinauer, Sunderland, MA.

Pimm, S. (1991). *The Balance of Nature*. Univ. of Chicago Press, Chicago.

Root, T. L., and Schneider, S. H. (1995). Ecology and climate: Research strategies and implications. *Science* 269, 334–341.

Schneider, S. H. (1997). *Laboratory Earth: The Planetary Gamble We Can't Afford to Lose*. Basic Books, New York.

Schneider, S. H., and Londer, R. (1984). *The Coevolution of Climate and Life*. Sierra Club Books, San Francisco.

Schneider, S. H., and Root, T. (1998). *Climate Change*. Reprinted from: Mac, M. J., Opler Puckettt Haecker, P. A., and Doran, P. D., Status and Trends of the Nation's Biological Resources. U.S. Department of the Interior, U.S. Geological Survey, Reston, VA. (This book includes all references cited in this article)

Torn, M. S., and Fried, S. J. (1992). Predicting the impacts of global warming on wildland fire. *Climate Change* 21, 257–274.

Trenberth, K. E. (Ed.) (1992). *Climate System Modeling*. Cambridge Univ. Press, Cambridge, UK.

Vegetation/Ecosystem Modeling and Analysis Project. (1995). Vegetation/Ecosystem Modeling and Analysis Project (VEMAP): Comparing biogeography and biogeochemistry models in a continental-scale study of terrestrial ecosystem responses to climate change and CO_2 doubling. *Global Biogeochem. Cycles* 9, 407–437.

Wilson, E. O. (1992). *The Diversity of Life*. Norton, New York.

Wright, H. E., Kutzbach, J. E., Webb, T. III, Ruddiman, W. E., Street-Perrott, F. A., and Bartlein, P. J. (Eds.) (1993). *Global Climates Since the Last Glacial Maximum*. Univ. of Minnesota Press, Minneapolis.

CLIMATE, EFFECTS OF

F. I. Woodward
University of Sheffield

GLOSSARY

net primary production Plant photosynthesis less plant respiration.

species diversity Species richness adjusted to account for abundance.

species richness The number of species recorded for a particular area, with each species given an equal weighting.

I. OBSERVATIONS

It is well established that the diversity of species increases toward the equator and that this observation is generally true for both animals and plants. The general term diversity indicates a measurement of the number of species which exist in a particular area. The measurement of species richness quantifies the number of species recorded for a particular area, with each species given an equal weighting. When species richness is adjusted to account for abundance, then species diversity is the appropriately termed measurement. Unfortunately, both measurements are often used interchangeably and with little precision. In this article, species richness is preferred, primarily because the aim is to investigate diversity at the global scale, when data are limited and, at best, may only include species lists, with some notional area of observation.

A. Species to Area Relationships

The area of ground within which species are counted plays a critical role in quantifying species richness. Rosenzweig (1995) provides an extensive account and examples of the relationships between species richness and area. One example is presented to illustrate the typical nature of the relationship for the flora of the British Isles (Fig. 1). Figure 1a shows the relationship for untransformed scales, with a particularly notable and rapid increase in species richness following small initial increments of area. Figure 1b presents the same data for log-transformed scales, in which the relationship is linear. Figure 1b indicates that the rapid increase in richness to about 1000 species occurs as the scale of observation increases from 0.1 ha to 1 million ha (10000 km^2 or 100 × 100 km).

All spatial scales contain data of ecological interest; however, in the context of the climatic control of diversity, the small areas are of less direct relevance. For example, the smallest scale in Fig. 1 of 0.1 ha (approximately 32 × 32 m) is too small to include many individual and mature trees. This is therefore a sampling prob-

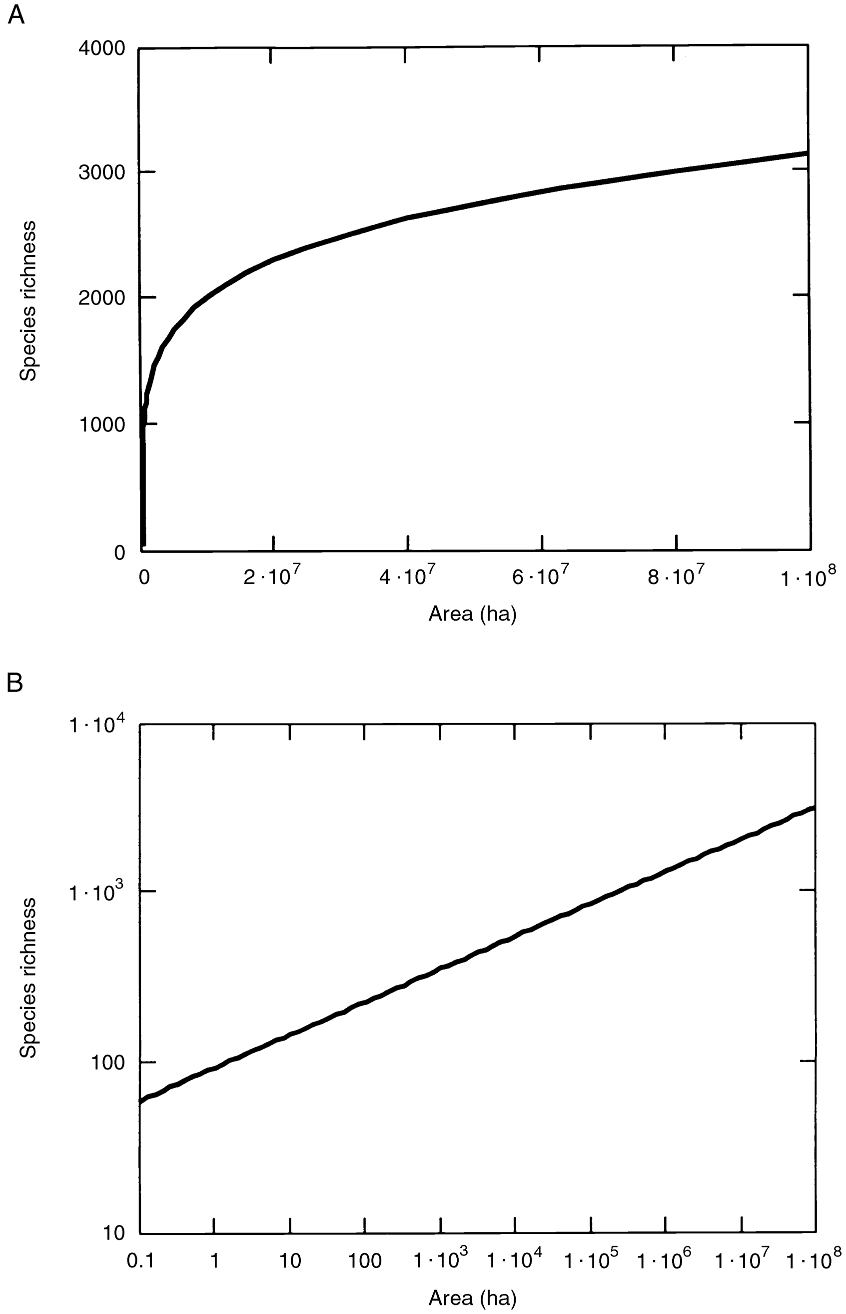


FIGURE 1 Species richness of the British Isles: (A) untransformed axes and (B) log-transformed axes (regression lines from Rosenzweig, 1995).

lem. The problem lies in deciding when this species to area relationship might indicate a primary control by climate, over and above any problems of sampling. It can also be envisaged that if an area under study has low landscape diversity (e.g., it is a flat plain of similar soil type), then this will also influence the species to area relationship. Such environmental features must be incorporated into general species lists and cannot easily be dissected out from the major thrust of this article—the influence of climate. A comparison of species richness with scale for the British Isles and the fynbos of South Africa (Fig. 2) indicates very large differences in diversity, but the untransformed scale (Fig. 2a) shows that the differences between the two floras decrease much more slowly above areas of about 2 × 10^7 ha (2 × 10^5 km^2 or 450 × 450 km). This indicates a useful order of scalar magnitude for comparing floras in the context of examining climatic control.

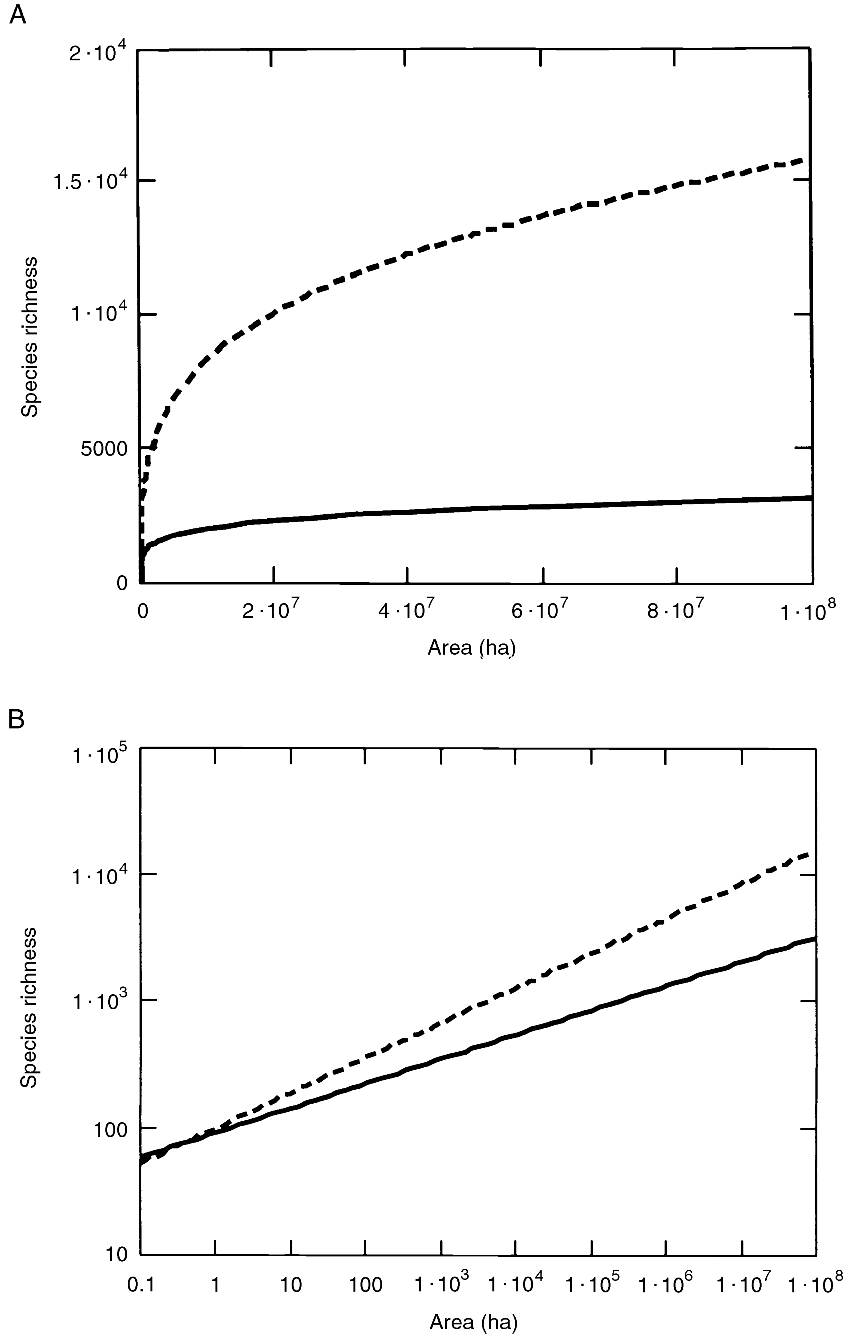


FIGURE 2 Species richness of the British Isles compared with the fynbos of South Africa: (A) untransformed axes and (B) log-transformed axes (regression lines from Rosenzweig, 1995).

B. Time Dependency

The species composition of an area endures changes with time, through mortality, disturbance, and migration. Disturbance of a moderate frequency can maximize species richness, producing landscapes in various stages of recovery from a disturbance event. An extreme disturbance event was the 1883 volcanic eruption which completely denuded the flora of the Krakatau Islands. A recent enumeration of higher plant species (Fig. 3; Whittaker *et al.*, 1990); showed a monotonic increase in the species complement during the past 100 years. It is clearly not possible to assess the final species richness of these islands until the species complement reaches some asymptote. This example of primary succession is perhaps extreme in the contemporary world, although smaller scale recoveries from natural disturbance and larger scale recoveries from human disturbance are current widespread features. Human disturbance is generally considered to deplete species richness and push natural vegetation to early stages of secondary succession; however, over the large spatial scales considered for this article it has been assumed that species richness has not yet been markedly depleted by human activities.

The last ice age was also the last natural and very large-scale disturbance to the terrestrial biosphere. At that time, glaciers spread over large areas of North America, Europe, and Russia and clearly removed all species in their paths. The cooler glacial climate, in addition to the spread of glaciers, led to the equatorial migration of plant species. However, in Europe, for example, the Alps prevented significant migration further south, with significant extinctions at this barrier. The climatic amelioration to the current interglacial climate reversed the process which occurred during the ice age. However, reduced richness follows as a consequence of the irreplaceable species extinctions. The degree of this loss of richness can be addressed in part from paleoecological information. However, such information fails to detect all species, and therefore it is most likely that glaciations cause some species losses and reductions in regional richness, but it is not possible to assess the extent of these effects. In any event, these long-term effects, and the shorter term effects of disturbance, are incorporated in regional assessments of species richness and, because they are at least partially climatically related, they emphasize the impacts of contemporary climate.

C. Observations by Area

The impacts of area on species enumeration and on the relationship between climate and species richness are well illustrated (Fig. 4) using data provided by Williams (1964);. Although the data are now outdated, in the sense that some larger areas are now better documented with greater richness, there is a clear distinction between areas which are wet and warm and those which are cold. Both the rate of increase in richness and the absolute richness with area are lower in the limiting, cold environment. Such observations conform with ob-

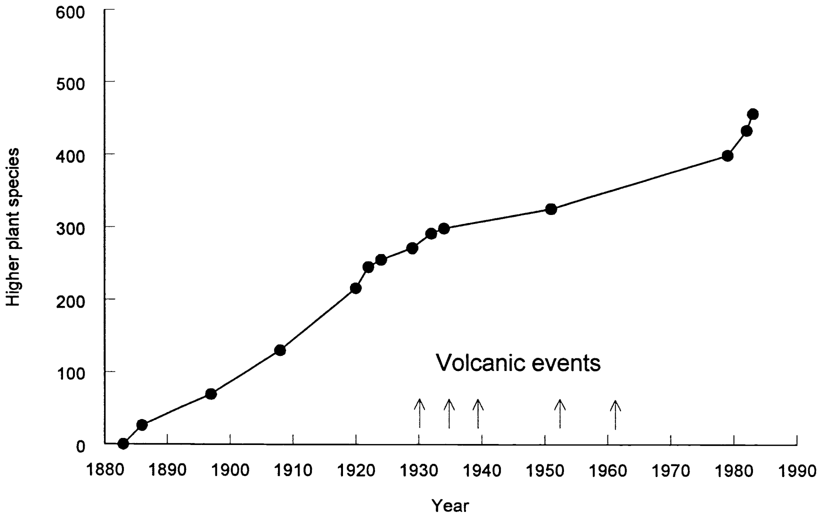


FIGURE 3 Measurements of species richness on the Krakatau Islands, Indonesia, since the major volcanic eruption of 1883. Subsequent and lesser volcanic events are also indicated by arrows (reproduced with permission from Whittaker *et al.*, 1990).

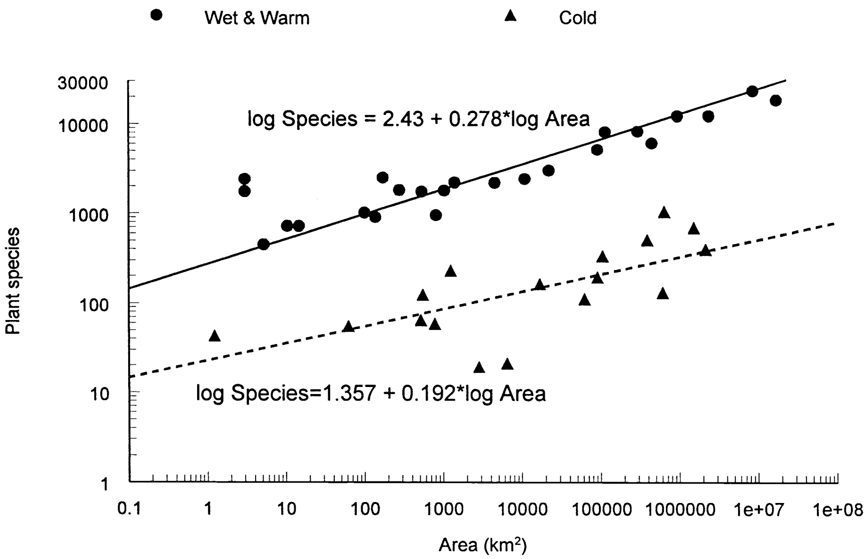


FIGURE 4 Species richness changes with area of observation and for areas with either wet and warm climates (●) or cold climates (▲), with significant regression coefficients and constants (data from Williams, 1964).

servations and modeling by Woodward (1987); which demonstrate a decreasing capacity for species survival with decreasing temperatures.

D. Global-Scale Observations

The data shown in Fig. 4 are restrictive in terms of global spread and accuracy, primarily because of the incomplete estimates of species richness at the continental scale. A more current global-scale distribution of species richness is available (Barthlott *et al.*, 1996;) on the Internet at the following address: *http://www.botanik.uni-bonn.de/system/biomaps.htm.* This map attempts to present plant species diversity data at the global scale, after allowing for variations in recording areas, using a globally uniform species–area relationship. This is a useful approach to circumventing the problems of unequal sampling areas, but it does hide the fact that species–area relationships will change with location and climate (Fig. 4). Nevertheless, this distribution map provides an important and unified source of data for assessing relationships between species richness and climate.

1. Species Richness and Latitude

The first cut at assessing species richness is usually obtained by latitude. This spatial variable includes many environmental features, particularly solar radiation receipt, day length, and temperature, but it is a useful variable for assessing broad-scale controls on vegetation. Figure 5 indicates a spread of richness with latitude. In the northern latitudes, the maximum observed richness decreases with increasing latitude. A similar relationship can be defined for the Southern Hemisphere, except that there are significant exceptions. These cases include the southeast coastal region of Brazil, extending from Porto Allegre (Fig. 5, Brazil 3) in the south through Rio de Janeiro (Fig. 5, Brazil 1) to Vitoria (Fig. 5, Brazil 2) in the North. Additional sites of high diversity have been sampled for South Africa (The Cape) and Australia (Perth). The sampling scheme for global diversity does not cover all areas of high diversity, but these five sites in the Southern Hemisphere and, to a lesser extent, the Mediterranean region in the Northern Hemisphere clearly indicate that areas of unusually high diversity do exist. These areas are defined as unusual because they seem outside the typical latitudinally dependent maximum extents of diversity.

2. Species Richness and Climate

a. Temperature

The observations of Fig. 4 suggest that cooler climates are less diverse. This same result is observed at the global scale (Fig. 6), when mean annual climate (for the 1970s and sampled at a scale of 10000 km^2) is compared with the Barthlott map of standardized species diversity (also adjusted to a 10000 km^2 scale). At

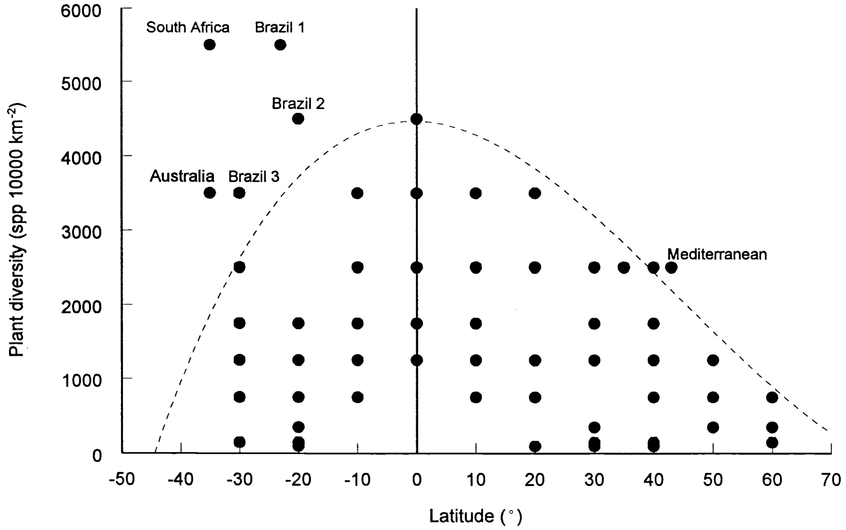


FIGURE 5 Changes in species richness with latitude. Details of the identified sites are discussed in the text (data from Barthlott *et al.*, 1996).

mean temperatures greater than about 5°C, species richness both increases markedly and becomes significantly more variable between individual sample points. The high-temperature, low-diversity sites are typically deserts or very arid, where it is clear that water supply is the critically limiting factor for plant survival. The areas of unusually high diversity in South Africa and Brazil are clear extremes in the temperature relationship, but the other sites shown in Fig. 5 are not distinct.

b. Precipitation

The relationship between richness and temperature is not simple and is also influenced by other climatic features, particularly precipitation. In contrast, 42% of the variance in richness (Fig. 7) is accounted for by a linear correlation between richness and annual precipitation. This association defines the obvious observation that diversity increases toward the tropics and is highest when warmth and high precipitation coincide. Four of the five recognized areas of high diversity remain as outliers from the remainder of the data set.

c. Net Primary Productivity

Precipitation is often taken as a surrogate of vegetation productivity in order to investigate any relationships between productivity, usually net primary productivity

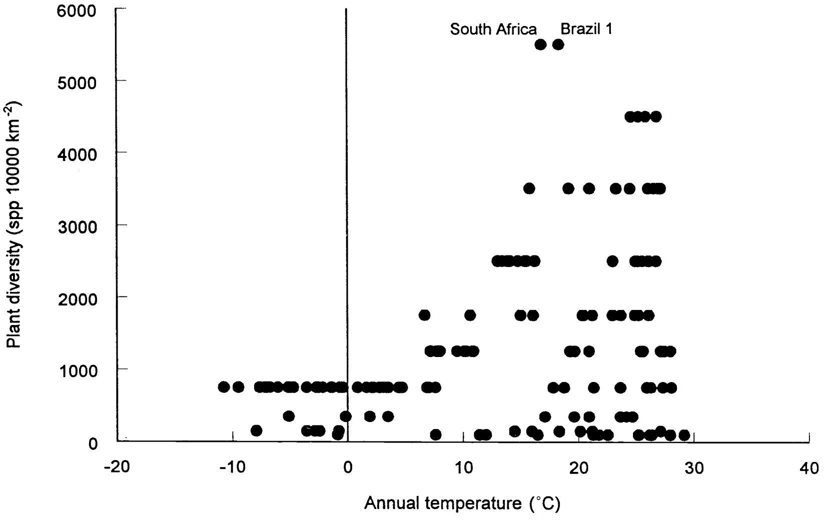


FIGURE 6 Changes in species richness with annual mean temperature.

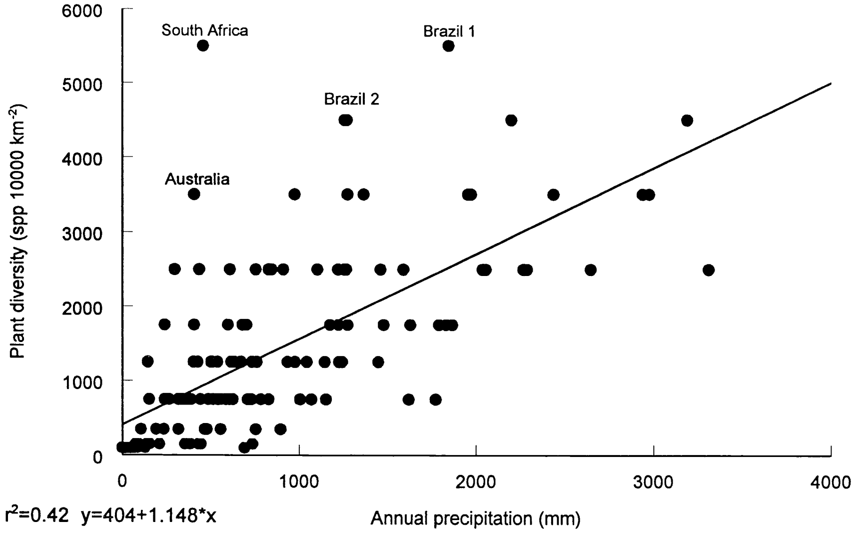


FIGURE 7 Changes in species richness with annual precipitation.

(NPP), and diversity or richness. Although precipitation exerts a large effect on NPP, this is not the only influence and other climatic features such as solar radiation receipt, day length, growing season temperature, and annual minimum temperature can exert significant effects on NPP (Woodward 1987). Simulation models of NPP typically incorporate the effects of all these climatic variables to provide an integrated effect, seen as simulated NPP (Woodward *et al.*, 1995).

In diversity studies, productivity is used and assessed in two different ways. For small field plot experiments, productivity is measured directly as the accumulation of biomass and is generally viewed as a consequence of diversity. On larger scales, intensive measurements of NPP are impossible and in this case productivity is generally simulated by some type of process model and it is more appropriate to consider diversity as a consequence of simulated NPP because this simulated value accounts for all variations in climate but is generally independent of any aspects of diversity.

In this case, a typical process model (Woodward *et al.*, 1995) has been used. Annual NPP is the net fixation of carbon into plant biomass, after plant respiration is decremented from annual photosynthetic fixation. In ecological terms, NPP is a measure of the rate of production of biomass (usually over an annual interval) and which is available for consumers and for creating new plant structure. Simulation models can often provide more accurate estimates of annual production, particularly in multiseasonal environments, which would require many harvests for accurate assessments of NPP in the field.

Simulated NPP provides a tighter fit than annual precipitation to the observed variations in species richness at the global scale (Fig. 8). Three of the high-diversity sites, however, are notable outliers from the line of best fit. The site in South Africa is the most extreme, with a very high richness associated with a low NPP. Much of this area of vegetation is known as *fynbos* and is a heathland on very impoverished soils. A further analysis of species richness in this area, and for the whole of southern Africa northwards to a latitude of 15°S, has been achieved by O'Brien *et al.* (1998). In this case, only the richness of woody species (native phanerophytes and plant heights of greater than 1.5 m) was considered. Woody species richness is very strongly correlated with NPP (Fig. 9) and the high-diversity site of Fig. 8 (arrow in Fig. 9) is not now distinguishable from the general trend. The high richness at this site is clearly due to herbaceous species and sub-shrubs and explanations for their high diversity may be related to factors not directly climatic, such as disturbance frequencies and types, the low nutrient status of the soils, and factors related to the history of the site, particularly rates of speciation and probabilities of species' migrations.

II. THEORIES

The strong latitudinal gradients in species richness have long been recognized and have led to a wide range of theories to explain the observations. At the same time, relevant new data sets are being continually produced

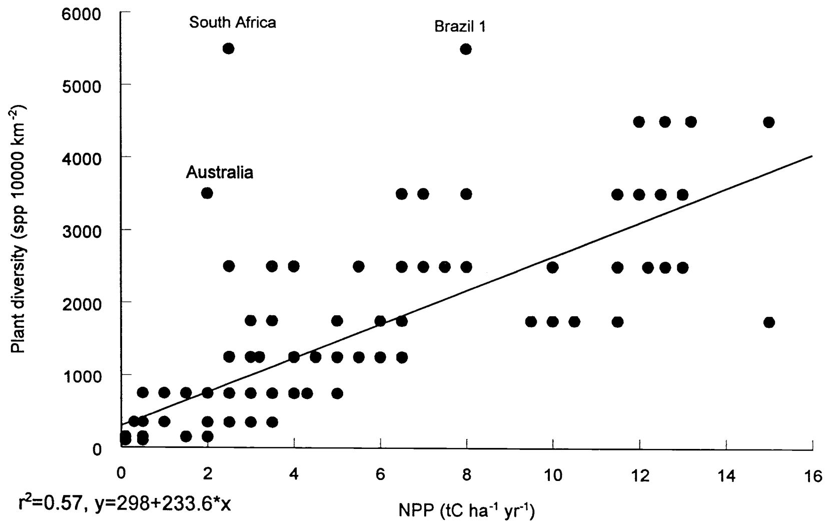


FIGURE 8 Changes in species richness with annual net primary production.

and analyzed in new ways. These data have been used to derive empirical or observational hypotheses about the controls of richness. In principle, if the empirical hypothesis is correct then it should also be in agreement with an appropriate theory. However, nature does not yield its secrets so easily and ambiguity seems to be the rule in that it is often difficult to differentiate between different theories and hypotheses when the differences in their consequences are often by degree rather than absolute. The following sections present current theories and hypotheses which address the relationships between climate and species richness.

A. Narrowing the Questions Using Observations

The defined area for observing species richness exerts a major impact in the quantity of richness. A less obvious, but also important, consideration is the nature of the areas considered at increasing scale. Figure 10

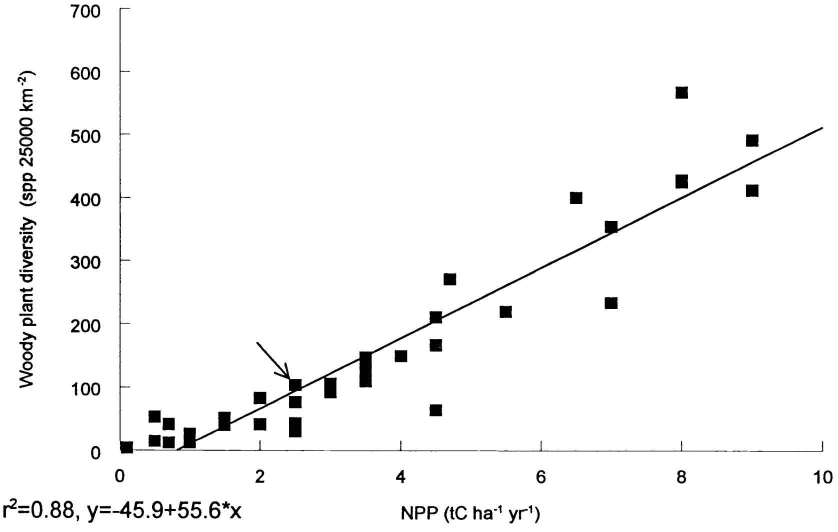


FIGURE 9 Changes in woody species richness of southern Africa with annual net primary production (NPP). The arrow indicates the NPP and woody plant richness for the site in South Africa indicated in Figs. 5–8 (data from O'Brien *et al.*, 1998).

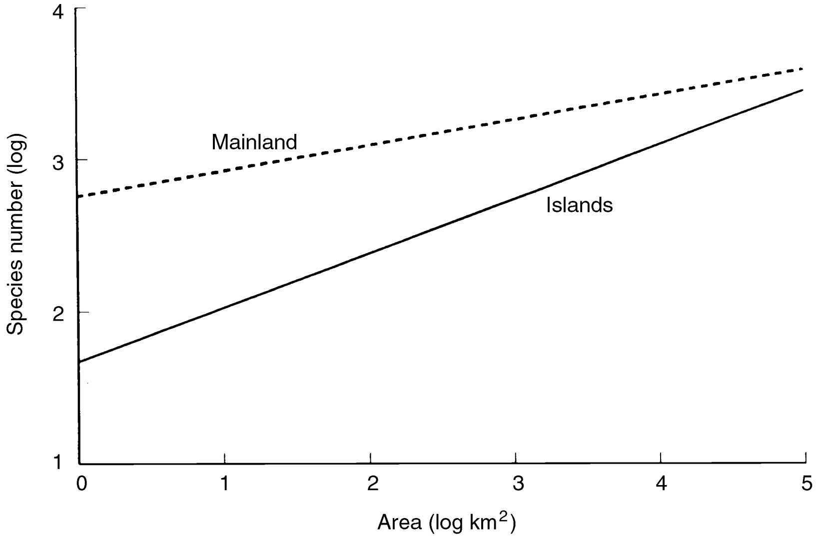


FIGURE 10 Changes in species richness with area for islands and mainland California (regression lines from Rosenzweig, 1995).

indicates that islands show a greater increase in richness with area than do areas of the same size but nested within a region with no dispersal and migratory barriers. The species richness of any area is the net result of species immigration, extinction, and evolution. For an island, the rate of species immigration will be less than that for a land bridge, and as a consequence the island richness will be less because local species extinctions are not readily matched by migration. As the islands become larger, the areas for the arrival of migrants and the survival of extant species increase and so extinctions will be less and migrations greater and, as a consequence, richness will increase so that it is similar to that of the nested land sites. This example shows that the nature of a land form and its degree of isolation from continental sites will influence richness in a manner which can be quite independent of climate, and therefore it is critical to account for such responses. An important contemporary corollary to the change in diversity with islands and mainland areas (Fig. 10) is that any process, such as human land clearance, which effectively makes small islands will inevitably lead to a loss of diversity.

The response shown in Fig. 10 also scales to larger areas (Fig. 11). In this case, the species to area relationship for the British Isles is much flatter than that for separate provinces and continents, from New Zealand, at the smallest, through Australia and Africa to the world. The flattest curve for the British Isles will include some component due to the net balance of extinctions and immigrations, plus a ready capacity for species to migrate throughout the British Isles, thus evening out, to some degree, the scale dependence of species richness. Different provinces and continents are much more isolated for immigration and also possess species complements determined by historical effects, not only past ice ages but also longer term speciation events. Climate will play and will have played a major part, therefore, in the contemporary species richness of these large areas. The climatic range of the British Isles, for example of temperature and precipitation, is small compared with that for a continent, but the range of soil conditions is wide from fertile alluvial soils to very low-nutrient bogs and chalk downs. Therefore, the species area curve (Fig. 11) for the British Isles also displays the influence of habitat diversity, particularly soil nutrient status on species richness, and the effect is small. It does not seem likely, therefore, that latitudinal gradients in soil nutrient conditions will play a major role in determining the gradients of richness seen, for example, with latitude (Fig. 5).

Observations by Rapoport (1982) may suggest a final constraint in understanding the climatic control of species richness. Rapoport noted the well-established observation that species richness declines with increasing latitude; however, this observation was supported by the observation that, on average, the geographical range of a species increased with latitude. Therefore, at high latitudes there are fewer species, but those which do occur are, on average, more widely dispersed. Subse-

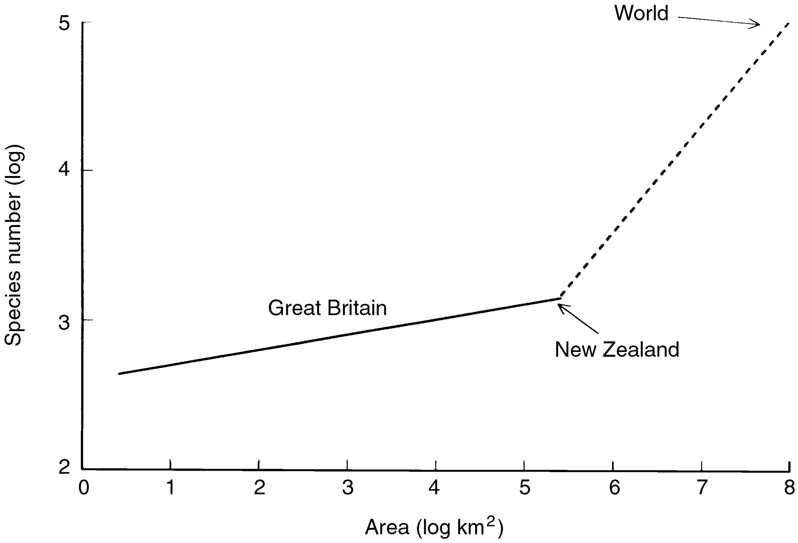


FIGURE 11 Changes in species richness with area within the British Isles (solid line) and between different floras (dashed line) (regression lines from Rosenzweig, 1995).

quent analyses (Gaston, 1999) have indicated that this relationship is neither universal nor simple. However, there is some consistency in the pattern from the tropics to high latitudes, particularly in the Northern Hemisphere.

B. Theories and Hypotheses

Theories and hypotheses for explaining the documented decline in diversity with latitude are based on one or more of the following fundamental controls:

- Competition for necessary resources
- Gap creation in vegetation by disturbance
- Energy availability
- Climatic extremes
- History

1. Competition

The competitive exclusion of species which require the same resource has been a theoretical underpinning to much of ecology, not just the climate control of diversity. Originating from simple experiments, it has proved easy to demonstrate the importance of competition in species interactions in response to changes in climate and in very small-scale experiments. It has not been so easy to demonstrate the importance of competition in defining the distributional limits of species on a larger field scale. Often, it seems that environmental heterogeneity in time and space can allow coexistence to occur, apparently circumventing direct competition for resources. However, competition is still considered an important mechanism in controlling survival and therefore diversity, but it does not appear to be easy to demonstrate competition in action.

It does seem that if competition is to play a major role in the control of species distributions, then it must do so as a secondary action to the primary loaction of climate. For example, consider the distribution of a boreal forest species in the cold of the high latitudes and a tropical species from the warmth of low latitudes. If competition plays a role in these distributions then it must favor boreal species at high latitudes and tropical species at low latitudes (i.e., competition is latitude dependent). Competition can only be latitude dependent through its impact on species' interactions and this would either be through a climatic dependency of species' contemporary or through historical responses. Therefore, it seems that competition is an integral and probably secondary part of a climatic control of diversity.

2. Disturbance

Creating gaps in vegetation by disturbance appears to provide a mechanism for avoiding direct competition and for providing bare ground and the opportunity for new species immigration. If some species are long-lived and can survive disturbance events, then they can co-occur with shorter-lived species which depend on disturbance and therefore diversity can be maintained. If the rate of disturbance changed systematically with latitude, then this could provide an explanation for the decline in diversity with latitude. However, Runkle (1989) has shown that this is not the case, with similar rates of disturbance with latitude.

Runkle (1989) did propose a climatic-dependent combination of disturbance with competition as a theory for controlling diversity. The length of the growing season decreases with latitude, which is a direct consequence of the curvature of the earth, a decreased annual incidence of solar radiation with latitude, and a consequent decrease in annual temperature. Plants in general can only grow when the temperature exceeds the freezing point of water, and therefore this constrains the period of growth in climates in which freezing temperatures occur. Because disturbance rates are broadly similar with latitude, it is expected that gaps in the vegetation will be created during the winter period of no growth. In the more uniform, or aseasonal tropical climates gaps will also be created through the year, but these will be filled rapidly because propagules are generally available throughout the year for establishment. In the seasonal climate, in contrast, gaps will accumulate until the temperature threshold for growth is exceeded. In the short growing season there is likely to be intense competition to establish in the early season in these gaps. These plants must also reach an appropriate stage of maturity and winter tolerance by the end of the growing season for survival, so there will be no survival advantage for delayed establishment. In this situation, competition can act to amplify the limiting effects of a short growing season.

3. Energy Availability

Hutchinson (1959) suggested that the availability of energy may constrain species diversity. Studies have since shown strong positive correlations between diversity and evapotranspiration, which is often taken as a surrogate for both energy and NPP. Figure 8 shows a clear relationship between NPP and species richness, but in this case climates with high energy inputs of solar radiation but low precipitation will lead to low NPP. O'Brien *et al.* (1998) describe the same response where water availability tempers a direct response of diversity to energy. Therefore, the relationship between diversity and energy is not simple, whereas the relationship between NPP and richness is in general quite linear. However, the mechanistic nature of the relationship is quite unclear in either case.

In some cases, particularly in small-scale experiments with variable resource supply, NPP and diversity are often negatively correlated, with just a small number of rank growing species dominating energy capture and production. Such a response is opposite to that observed at the larger, global scale (Fig. 8) and, based on Fig. 1a, indicates that climatic control of diversity is difficult to determine when the area of observation is on the steeply rising curve of the species–area relationship. Nevertheless, small-scale experiments have often influenced ideas at the global scale. It is generally considered that sites of high NPP are highly competitive with only a small number of dominant species. However, highly productive tropical forests are also highly diverse in terms of both species and vegetation structure, even with limiting soil nutrient resources. The vegetation canopy may have many strata, with different species in each strata. High NPP requires a high supply rate of resources, particularly soil nutrients, water, and solar radiation, with no significant temperature limitations. On low-nutrient soils, the vegetation retains the majority of the soil nutrients, whereas water is in abundant supply. In these forests, turnover of plant parts (leaves, twigs, and small branches) is quite rapid, as is the decomposer system and thus also the rate of nutrient recycling. Therefore, the high NPP goes hand in hand with a high requirement and supply of nutrients to all component species. This appears to be the major mechanism by which high NPP can sustain high diversity. In addition, large areas of forest also transpire large quantities of water vapor, which returns as rain and also increases local cloudiness. The high NPP vegetation therefore exerts some control over its own climate, making it more equable and reducing energy supply.

4. Climatic Extremes

Climate appears to influence diversity by a range of mechanisms, such as by influencing turnover times of soil nutrients and through constraining the length of the growing season. However, it is also well-established that species have a range of minimum temperature tolerances which is much greater than the range of maximum temperature tolerances (Woodward, 1987). At the global scale, absolute minimum temperatures in vegetated areas of the world vary from about −80 to 20°C and species are appropriately adapted to the local absolute minimum temperature, even if this occurs only every 10–30 years. Therefore, minimum temperature tolerance increases with latitude. There also appears to be a cost involved in possessing a particular degree of temperature tolerance, with the cost probably reducing growth capacity and competitive ability compared with those of species possessing a lower minimum temperature tolerance (Woodward, 1987). As the absolute minimum temperature decreases, so also does the number of species that can survive (Woodward, 1987).

The absolute minimum temperature decreases with latitude; therefore, this mechanism alone could account for the latitudinal trend in species richness (Fig. 5) if fewer species have evolved the capacity to endure increasingly lower temperatures. However, this does not account for the absence of low-temperature-tolerant species from warm climates; in addition, low absolute temperatures are also associated with shorter and cooler growing seasons. Therefore, the impact of climatic extremes on richness almost certainly occurs in parallel with other climatic limits, such as described by Runkle (1989; see Section II.B.2).

It is possible to use simple models to investigate the impacts of changes in absolute minimum temperature and competition on distribution and diversity. Figure 12 shows a typical output of a model in which species are designed with the capacity to survive to pre-determined minimum temperatures but have the capacity to survive at all higher temperatures. In a second case, the competitive ability of the species declines as temperatures increase above the low temperature cutoff for survival. In both cases of competition and no competition, richness increases with temperature. However, in the simulation with competition diversity increases less at temperatures above −20°. This occurs because competition excludes those species with the lowest temperature tolerances from the highest temperatures. This does not occur in the simulation with no competition. Figure 13 illustrates the case for a species with a minimum temperature tolerance to −35°C. Competition exerts no effect on abundance to −30°C but at higher temperatures the abundance or occurrence of this species decreases consistently to zero at −5°C. Without competition there is still a reduction in the occurrence of this species with increasing temperature, and this occurs simply because the chance of occurrence decreases as the pool of species increases with temperature. Figure 12 demonstrates that richness decreases with decreasing temperature; however, the occurrences of the tolerant species increase, in line with the suggestion by Stevens (1989).

The model is simple but it readily predicts changes in diversity with temperature and also indicates that a reduction in richness will be associated with an increase in the abundance of species. Both examples (Figs. 12 and 13) also indicate that the effect of competition is primarily of degree, and therefore it will prove very difficult to differentiate between direct competitive effects and a simple lottery model as both temperature and the species pool increase.

5. History

It appears that competition, disturbance, NPP, and climatic extremes can all influence the way in which climate impacts on diversity. It also seems likely that we

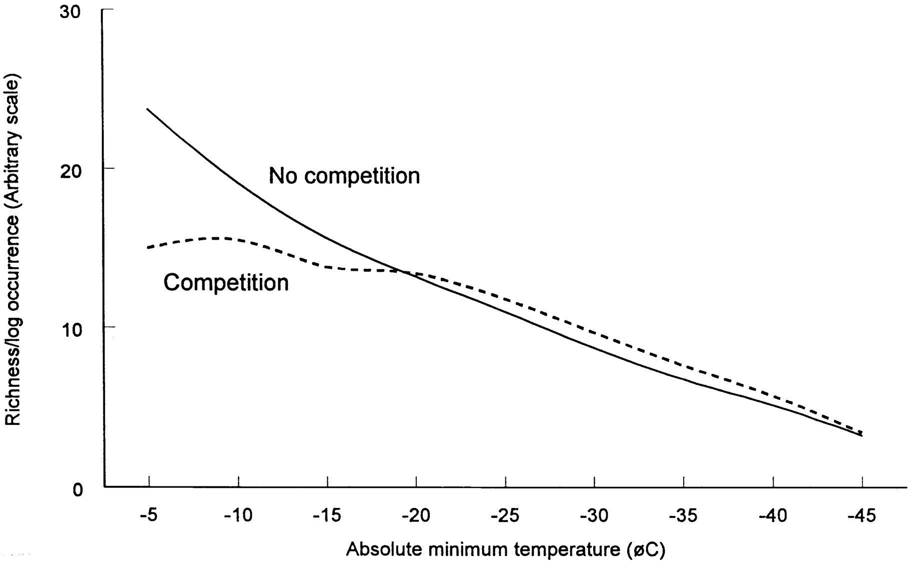


FIGURE 12 Simulated changes in species richness with absolute minimum temperature, without competition (solid line) and with competition (dashed line).

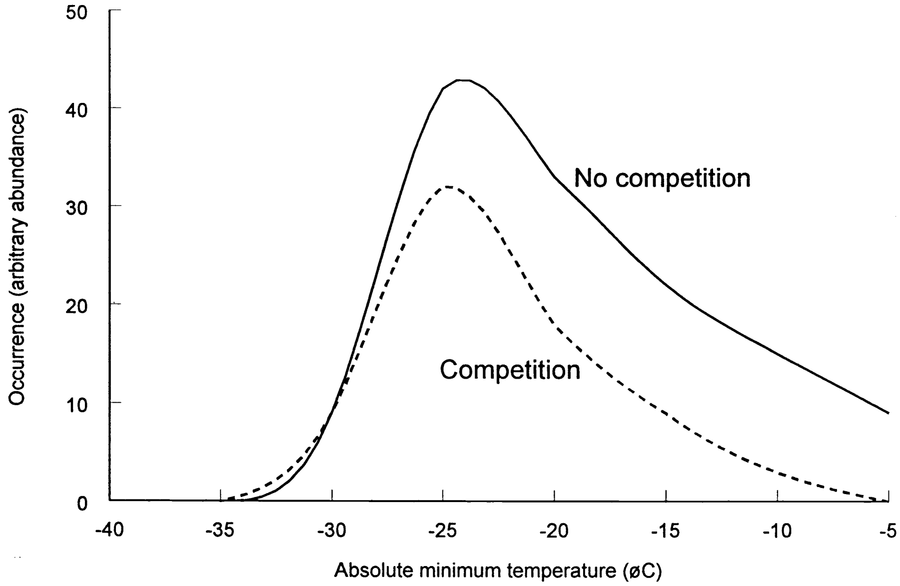


FIGURE 13 Simulated occurrence of a species with a minimum temperature tolerance to −35°C, without competition (solid line) and with competition (dashed line).

cannot readily differentiate between these controls at the global scale. However, none of the models are capable of explaining the sites which are outliers in the general relationships between richness and latitude (Fig. 5), temperature (Fig. 6), precipitation (Fig. 7) and NPP (Fig. 8). Further analysis indicates that these areas all have high frequencies of endemic species and at the global scale species richness is positively correlated with the degree of endemism. The high degree of endemism in the fynbos of South Africa and kwongan of South West Australia appears to be strongly correlated with the degree of soil infertility. In these areas, richness decreases with increasing soil fertility. Both landscapes, however, are very infertile and ancient in geological terms, and it seems likely that the infertile soil limits the capacity both for alien species to invade and for greater genetic mixing and consequent lower levels of endemism than currently occur. The coastal Atlantic forests of Brazil are isolated from other areas by an inland mountain chain, indicating another method of limiting migration and cross-fertilization and enhancing species richness.

Processes associated with history and regional landscape structure seem to be the most parsimonious explanations for the high species richness in these areas. In the Cape floral kingdom of South Africa, there are more than 500 species in the genus *Erica* (Bond, 1989). The species are structurally very similar and it seems unlikely that their ecological characteristics are very different. Such situations suggest that the species concept is not an ideal measure of richness, particularly ecological richness. The indicated areas of high diversity (Fig. 5) certainly indicate high taxonomic richness, but it appears that this is hardly relevant to the climatic controls of richness.

III. EXPERIMENTS

Numerous experiments have investigated the relationship between species richness and productivity. However, through practical necessity these experiments have been carried out on a small scale, well within the range of rapidly increasing richness with area as shown in Fig. 1. This scale is not appropriate for the larger scale considerations discussed in this article. Indeed, Tilman (1999), a major contributor to the productivity–diversity debate with Grime (1997), has indicated that this understanding, developed from scales where sites are populated from the same regional species pool, is inappropriate for addressing larger scale latitudinal patterns.

One experimental approach which can, in part, be used to address the relationship between diversity and climate is that of a herbaceous productivity–diversity multi-site experiment distributed across Europe (Hector *et al.*, 1999). Eight sites were distributed between 39° and 64°N of latitude and 8°W and 27°E of longitude.

In all cases, these grassland experiments were seeded from bare with a known number of species, from one to at least eight. At all sites, aboveground biomass, a measure of productivity, increased with species richness, in keeping with the correlation shown in Fig. 8. Although the expected reduction in biomass with short growing season (either through drought or low temperatures) was clearly demonstrated, it was not possible, in a 2-year experiment, to determine the effects of climate on the equilibrium species richness of the plots.

A. Genome Size

In addition to resolving the mechanisms by which climate controls species richness, it is also important to be predictive. For example, given a particular climate, what will be the diversity and what will be the characteristic behavior of the local species? An alternative question is, given a particular metric of a species, where might the species be distributed naturally and in how diverse a mixture of species? These questions cannot be resolved; however, there has been a continued interest in relating the size of the nuclear genome of a species in this way (Grime, 1998). Relationships between genome size and critical climatic tolerances of frost resistance, precipitation supply, and growing season have been found, as have relationships between genome size and plant growth dynamics. This is still clearly an area for further development, but in many ways the attraction to the approach is not the current certainty of its predictive skills but rather the way in which it allows questions to be asked and refuted or otherwise at the local to global scales.

IV. CONCLUSIONS

The impacts of climate on plant biodiversity are best studied on the large scale, in the order of hundreds of kilometers. This approach escapes the insoluble problems of disentangling the effects of scale, landscape structure, and negative small-scale relationships between productivity and diversity. Freed from these constraints, the large-scale nature of the study indicates that observations and correlations are the primary mechanisms for understanding the processes by which climate controls diversity. The processes are still not fully clear, but primary productivity, disturbance, the length of the growing season, and competition all appear to play important roles. There are some locations where species richness is much larger than the average. These areas are rich in endemic species which have probably originated and persisted because of migratory and reproductive barriers. In these areas, it is likely that species differ very little or at all in ecological characteristics, and in the context of this article such diversity, over and above the average, has little meaning.

See Also the Following Articles

CLIMATE CHANGE AND ECOLOGY, SYNERGISM OF • COMPETITION, INTERSPECIFIC • DISTURBANCE, MECHANISMS OF • ENERGY FLOW AND ECOSYSTEMS • LATITUDE, COMMON TRENDS WITHIN • SPECIES-AREA RELATIONSHIPS

Bibliography

Barthlott, W., Lauer, W., and Placke, A. (1996). Global distribution of species diversity in vascular plants: Towards a world map of phytodiversity. *Erdkunde* **50**, 317–327.

Bond, W. J. (1989). Describing and conserving biotic diversity. In *Biotic Diversity in Southern Africa: Concepts and Conservation* (B. J. Huntley, Ed.), pp. 2–18. Oxford Univ. Press, Cape Town.

Gaston, K. J. (1999). Why Rapoport's rule does not generalise. *Oikos* **84**, 309–312.

Grime, J. P. (1997). Biodiversity and ecosystem function: The debate deepens. *Science* **277**, 1260–1261.

Grime, J. P. (1998). Plant classification for ecological purposes: Is there a role for genome size? *Ann. Bot.* **82**(Suppl. A), 117–120.

Hector, A. et al. (1999). Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127.

Hutchinson, G. E. (1959). Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* **93**, 145–159.

O'Brien, E. M. Whittaker, R. J. and Field, R. (1998). Climate and woody plant diversity in southern Africa: Relationships at species, genus and family levels. *Ecography* **21**, 495–509.

Rapoport, E. H. (1982). *Aerography: Geographical Strategies of Species*. Pergamon, Oxford.

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time*. Cambridge Univ. Press, Cambridge, UK.

Runkle, J. R. (1989). Synchrony of regeneration, gaps, and latitudinal differences in tree species diversity. *Ecology* **70**, 546–547.

Stevens, G. C. (1989). The latitudinal gradient in geographical range: How so many species coexist in the tropics. *Am. Nat.* **133**, 240–256.

Tilman, D. (1999). The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* **80**, 1455–1474.

Whittaker, R. J., Bush, M. B., and Richards, K. (1990). Plant recolonization and vegetation succession on the Krakatau Islands, Indonesia. *Ecol. Monogr.* **59**, 59–123.

Williams, C. B. (1964). *Patterns in the Balance of Nature*. Academic Press, London.

Woodward, F. I. (1987). *Climate and Plant Distribution*. Cambridge Univ. Press, Cambridge, UK.

Woodward, F. I., Smith, T. M., and Emanuel, W. R. (1995). A global primary productivity and phytogeography model. *Global Biogeochem. Cycles* **9**, 471–490.

COASTAL BEACH ECOSYSTEMS

Anton McLachlan
College of Science, Sultan Qaboos University, Oman

GLOSSARY

accretionary Accumulating sand.

dissipative beaches Wide, flat beaches.

effluent line Level on the beach face where the water table intersects the surface.

macrotidal Describing a spring tide range of more than 4 m.

microtidal Describing a spring tide range of less than 2 m.

morphodynamics Interactions between the physical structure and the water and sediment movement in a beach and surf zone environment.

psammophilic Sand-loving.

reflective beaches Narrow, steep beaches.

transect A linear series of samples, for example, across the intertidal zone.

supralittoral Immediately above the intertidal zone.

EXPOSED OCEAN BEACHES ARE HARSH AND DYNAMIC ENVIRONMENTS controlled by waves, tides, and sediment type. Global patterns of macrofauna biodiversity on sandy beaches are tightly coupled to these physical features of the beach environment and can be reliably predicted on the basis of beach type.

I. INTRODUCTION

The world has entered a biodiversity crisis. The rapid loss of species as a consequence of human alteration of the biosphere has attracted widespread attention among ecologists and spawned a voluminous literature concerning the cataloguing, description, prediction, rescue, and management of biodiversity. Marine ecologists have long been interested in explaining patterns in community diversity, especially in the benthos. Much of this effort has been directed at deciphering the role of biological interactions, such as competition, in maintaining diverse communities, and this debate has perhaps been most pronounced amongst workers interested in deep sea-floor and rocky shore environments. However, it has been suggested that competition is relatively unimportant in intertidal sediments (Peterson, 1979), where predation, recruitment, and other processes may exert greater influence on community structure (Reise, 1985).

Despite implicit assumptions of the underlying importance of physical environmental features and processes in influencing diversity in marine benthic communities, it has been much more fashionable to look for biological explanations, especially in competitive interactions. This has perhaps been an oversight and abiotic factors have not been given adequate attention. This article addresses this shortcoming by examining a physically controlled ecosystem and determining to what extent biodiversity can be predicted by physical variables alone.

Sandy beaches dominate the ocean shorelines of all temperate and tropical continental coasts (Davies, 1972). Sandy beaches are devoid of any biological structures and their morphology and dynamics can be defined in terms of three interacting factors: waves, tides, and sand particle size. This simplicity should make understanding of such systems relatively elementary. However, relative to other shore types, beaches have been rather neglected by ecologists.

Here I describe the global range of wave-exposed, sandy beach types that can occur in response to changes in the values of their three defining variables, namely, waves, tides, and sand. I explore the extent to which the diversity and abundance of their intertidal macrobenthic communities can be explained by these factors, and then consider biological and biogeographic issues.

II. SANDY BEACH TYPES

Ocean beaches are defined by the interactions of the wave energy they experience, their tidal regimes, and the nature of the sand available for sorting and transport by the tides and waves. The simplest overall index of beach state is the beach slope, which is a product of the interaction among all three of these variables (Bascom, 1980; Short, 1996): beach face slopes flatten as wave energy increases, tide range increases, or particle size decreases, if other factors are kept constant. Thus the flattest beaches occur in macrotidal regions of high wave energy and fine sand, and the steepest beaches occur in microtidal regions with low wave energy and coarse sand. A range of beach morphodynamic types can be distinguished between these extremes.

In a microtidal regime, where beaches are wave dominated, three beach states can be recognized: reflective, intermediate, and dissipative. The reflective beach, characterized by a steep face and absence of a surf zone, occurs under a combination of coarse sand and gentle waves. The shoreward transport of sand, which occurs under these conditions, causes all sediment to be stored on the subaerial beach face; the reflective beach thus represents the accretionary extreme in beach states. Waves surge up the beach face, where they may break before being reflected back to sea (Fig. 1).

Dissipative beaches, in contrast, are a product of large waves moving over fine sand. This results in a flat beach face and wide surf zone. Waves break far out and dissipate their energy while traversing the surf zone as bores before expiring as swash on the beach face. Dissipative beaches, with their sand spread out over extensive surf zones, thus represent the erosional extreme in beach states. Between these two extremes, intermediate beaches are distinguished by the presence of surf zones that are smaller than in the dissipative situation and generally 20–100 m wide. The intermediate surf zone characteristically has well-developed bars (sandbanks) and channels with rip currents (see Fig. 1).

Beaches are not locked into a single morphodynamic state and respond to changes in wave energy by moving towards dissipative conditions during storms (and spring tides, which are of maximum range and occur during the new and full moons) and towards reflective conditions during calm weather (and neap tides, which are of minimum range and occur during the first and third quarters of the moon); that is, sand erodes or accretes on the beach face as wave height (and tide range) rises or drops.

There is a useful index that conveniently describes the state of a microtidal beach, that is, the extent to which wave energy is dissipated or reflected. Dean's parameter, also known as the dimensionless fall velocity, is given by:

$$\text{omega} \equiv \text{wave energy/sand fall velocity}$$

where wave energy is given by modal breaker height (cm) divided by modal wave period (seconds) and sand fall velocity is the sinking rate (cm per second) of the mean sand particle size on the beach. Values for omega that are <2 generally indicate reflective beaches and values >5 indicate dissipative beaches.

The foregoing description of beach types adequately covers most microtidal situations, but increasing tide range complicates the picture and requires further explanation. Tides play a role essentially similar to that of waves in that increasing tide range tends to make beaches even more dissipative (Short, 1996). This occurs because increasing tide range allows the surf zone to work back and forth over a wider area. Indeed, fully reflective beaches will not occur when tide range exceeds 1–1.5 m. On beaches with larger tides reflective conditions can only occur at the top of the shore between the neap and spring high-water swash lines, an area reached by swash but not surf—this area is con-

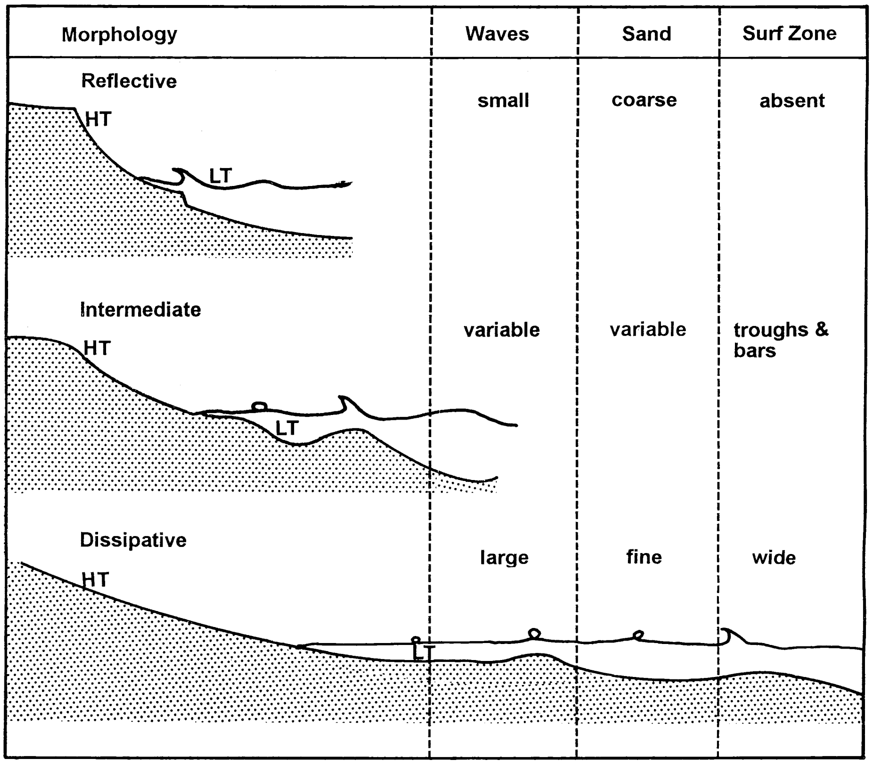


FIGURE 1 Three morphodynamic states of microtidal beaches (HT = high tide, LT = low tide).

trolled by swash processes and is accretionary and steep, in contrast to the rest of the shore, where shoalling and breaking wave and bore processes operate.

Under large tidal regimes (mean spring range >4 m) beaches are generally tide dominated, whereas in intermediate situations (tide range 2–4 m) they are mixed and either waves or tides can dominate. A useful index of the relative importance of waves and tides is the relative tide range (RTR), which is given by the mean spring tide range divided by the modal breaker height. By combining the dimensionless fall velocity and RTR, a two-dimensional model of beach states can be produced (Fig. 2).

Besides beach slope, omega, and RTR, another useful index of beach morphodynamic conditions is the beach state index (BSI; McLachlan *et al.*, 1993). BSI combines measures of wave energy, tide range, and sand fall velocity into one index:

$$\text{BSI} = \log[(\text{omega}) \times (\text{maximum tide range}/0.8\text{ m})]$$

This index is rendered dimensionless by dividing the tide range by an equilibrium tide range of 0.8 m. The BSI has values between 0 and 2; microtidal, reflective beaches score <0.5, intermediate to dissipative beaches score 0.7–1.5, and macrotidal ultradissipative beaches and sand flats generally score >1.5.

III. SAMPLING BEACH MACROFAUNA

Community studies of beach macrofauna make use of standard beach transect surveys. These typically involve quantitative sampling across the intertidal zone by excavating quadrats and passing the sand through a screen of 1-mm mesh, the optimum mesh size. Other mesh sizes, for example, 0.5 mm or 2 mm, have also been used and can significantly influence the results. The finer mesh size is more effective in sampling larval stages of macrofauna and can collect some larger meiofauna, but it is not practical in coarser sands.

In the context of this article, each transect survey represents one datum point in the sense that beaches are compared as if they were units. Data should only be used from transect surveys where the total sampling effort (assessed as total area excavated, i.e., levels ×

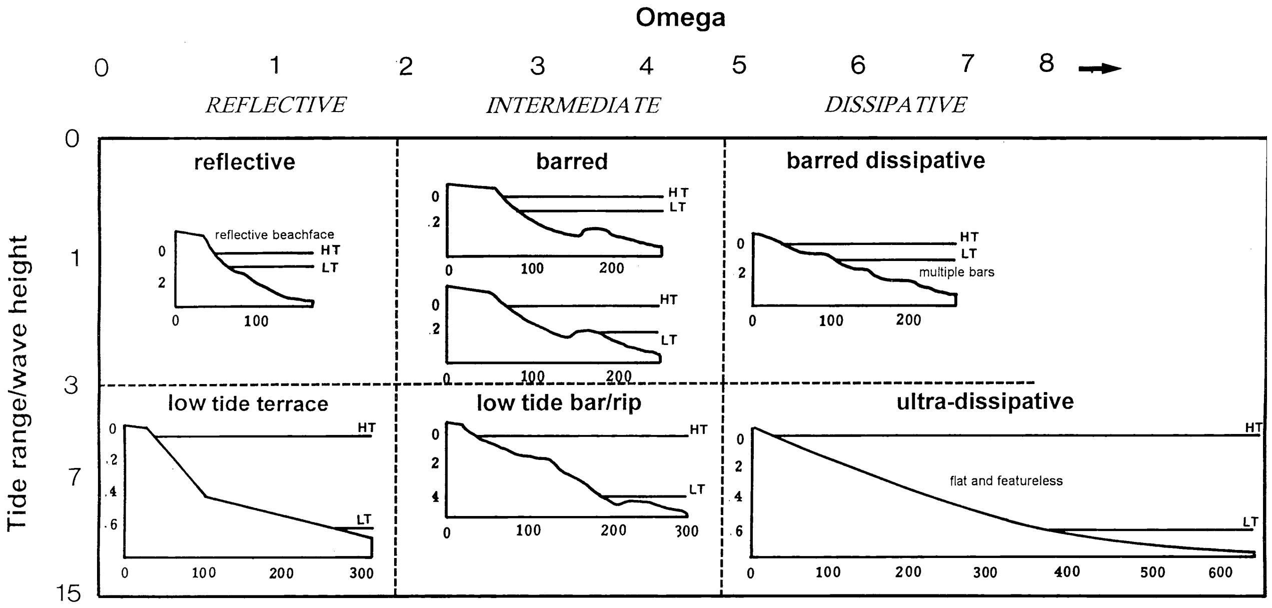


FIGURE 2 Two-dimensional model of beach states. (After Short, 1996.) Distances are in meters.

replicates × quadrat size) per beach exceeded 2 m^2. Smaller sample sizes introduce problems of undersampling, as illustrated by species/area relationships in Fig. 3 (Jaramillo *et al.*, 1995). The absence of an asymptote in Fig. 3 means that total species richness is never sampled, but the curves indicate that, for microtidal beaches a sample area of 3 m^2 is effective; whereas macrotidal beaches require more. I will use richness rather than diversity indices because it is more conservative. Species richness (or number of species) in beach surveys is summed for all samples in a transect and abundance is usually calculated per running meter of transect, so results are expressed as the number of species per transect and the number of individuals per linear meter of transect.

IV. COMPOSITION AND ZONATION

The intertidal macrofauna of ocean sandy beaches is usually dominated by crustaceans, molluscs, and polychaetes, with other groups, such as insects, nemertean worms, echinoderms, anemones, and fishes, being of minor importance or restricted to the extreme upper or lower fringes. Crustaceans tend to be most successful in reflective conditions, where their great mobility enables them to cope with turbulence. Among the crustaceans, ocypodid crabs, hippid crabs, cirolanid isopods, a variety of amphipods, and psammophilic mysids are

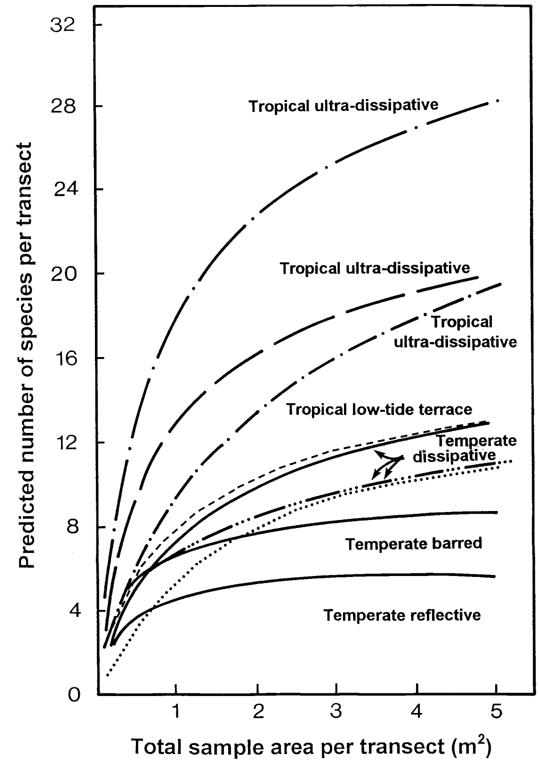


FIGURE 3 Species/area curves for sandy beach macrofauna. (After Jaramillo *et al.*, 1995.)

the most typical. Molluscs, both gastropods and bivalves, are successful over a wide range of beaches. Clams, in particular, may form large populations on high-energy dissipative beaches, where they can support commercial fisheries. Polychaetes tend to be the group that is most sensitive to beach state and are absent or scarce on reflective or coarse sand beaches. On lower-energy shores of fine sand, polychaetes can be particularly abundant and include predator/scavengers and deposit and suspension feeders.

Like all intertidal regions, sandy beaches display zonation of their macrofauna. Various models have been discussed, but the most applicable is that of Erik Dahl (see McLachlan and Jaramillo, 1995). This model identifies three zones on sandy beaches: a supralittoral zone characterized by ocypodid crabs in warm regions and talitrid amphipods in temperate areas; a littoral or midshore zone characterized by cirolanid isopods and spionid and opheliid polychaetes; and a lower shore or sublittoral fringe with many groups, including hippid crabs, mysids, haustoriid and phoxocephalid amphipods, donacid clams, and nephtyid worms. The lower zone tends to reduce or even disappear in very reflective situations, but it can expand to a broad, species-rich terrace under dissipative conditions. The supralittoral zone, abovethe high-water mark or drift line, is present on all shores. Thus, on most shores there is a clear gradient of increasing species richness as one moves downshore, typically from one or two species in the supralittoral to many species on the lower shore. This article focuses on between-beach diversity rather than variability within beaches.

V. GENERAL PATTERNS

McLachlan *et al.* (1981) were the first to demonstrate a relationship between beach type and biodiversity. Based on a survey of beaches around South Africa, they showed that the species richness of benthic macrofauna increased as beaches became flatter and particle size became finer (Fig. 4). They also demonstrated that faunal abundance increased exponentially as particle size decreased and beaches became flatter. At that time they were not aware of the beach state models that were first brought to the attention of ecologists two years later (Short and Wright in McLachlan and Erasmus, 1983). Since the South African coast is uniformly microtidal and subject to vigorous wave action throughout, these findings largely reflect the role of particle size in influencing microtidal beach state. The importance of beach slope, sand grain size, and wave exposure has also been stressed by other authors.

After taking into account the significance of morphodynamic models, a subsequent study that examined a wider range of beaches clearly showed a linear increase in species richness (and an exponential increase in abundance) per transect over a range of beach types from reflective to dissipative (Fig. 5; McLachlan, 1990). Taking this further, McLachlan *et al.* (1993) showed this trend to be global, based on data from four conti-

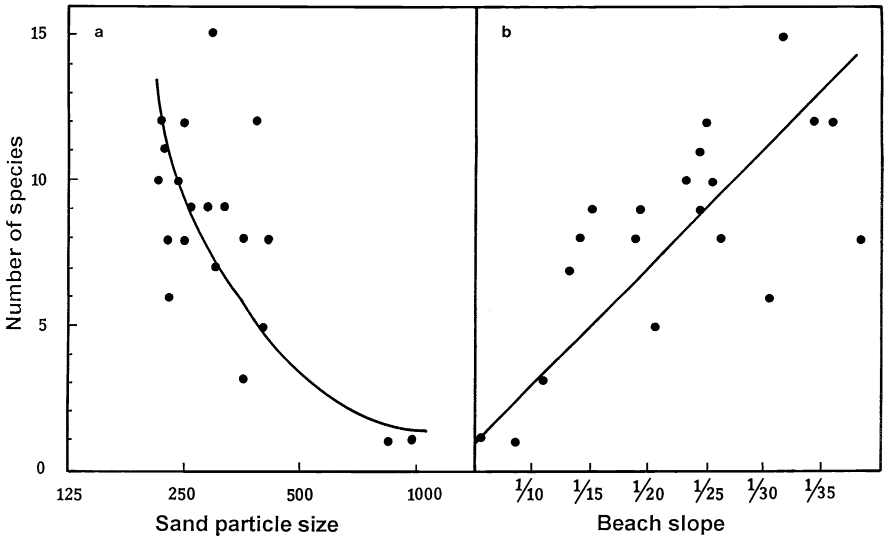


FIGURE 4 Species richness related to beach slope and grain size (in microns); early data from South Africa. (After McLachlan *et al.*, 1981.)

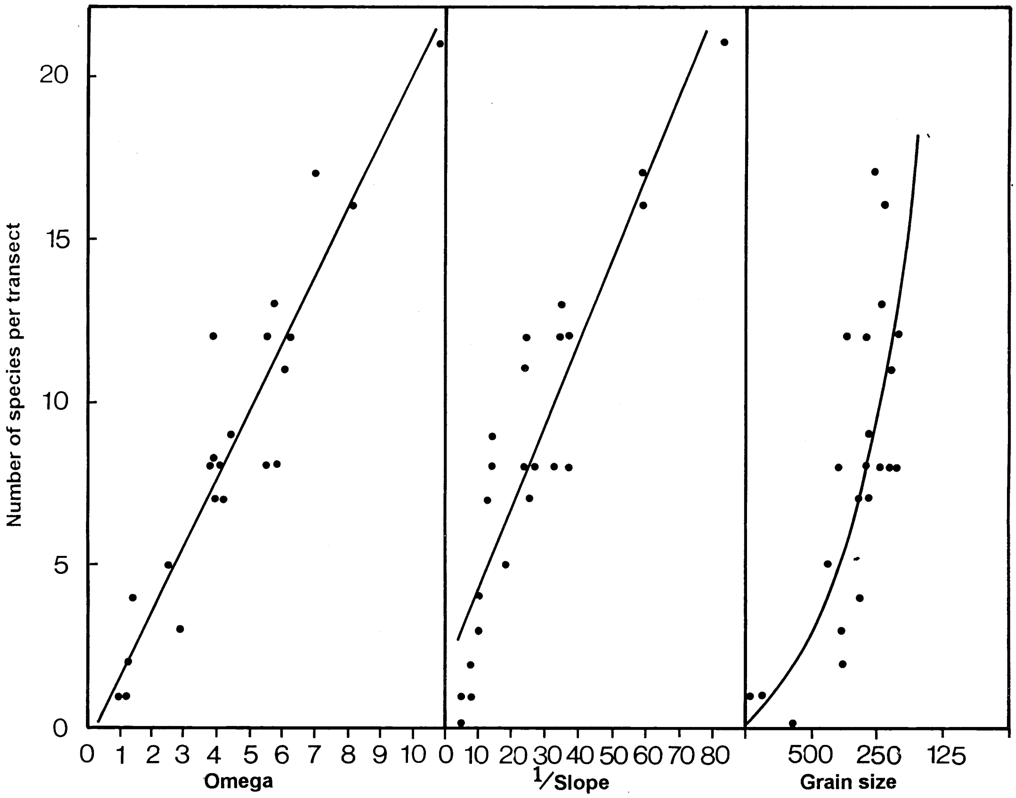


FIGURE 5 The first analysis of global patterns of species richness on sandy beaches. Grain size is in microns. (After McLachlan, 1990.)

nents (Fig. 6), and McLachlan *et al.* (1996) then examined an even wider data set, including macrotidal beaches, and showed excellent correlation using the BSI index. Further confirmation has come from data sets from Australia (Fig. 7; Hacking, 1997), New Zealand, Brazil, Madagascar, Spain, the United States and Oman.

The foregoing studies have confirmed that, for a diverse range of exposed sandy beach environments on five continents, intertidal macrofauna species richness per transect increases linearly with increasing values of BSI. Thus the number of species accommodated on an intertidal beach increases in response to increasing tide range, increasing wave energy, and decreasing sand particle size. In simple terms, this means that species richness increases as beaches become flatter and wider.

The width of sandy beaches ranges from as little as 10 m in reflective, microtidal situations to hundreds of meters in dissipative tidal flats, thus representing an order of magnitude range in the length of an intertidal transect. The total area sampled by various authors, however, has fallen within a much smaller range. Total

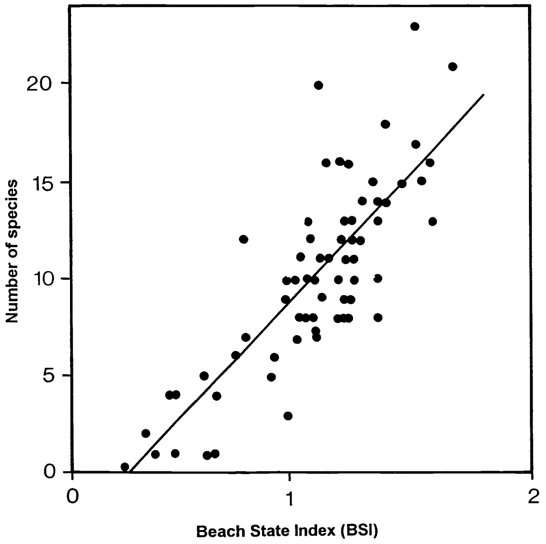


FIGURE 6 The first plot of species richness against BSI, based on data from four continents. (After McLachlan *et al.*, 1993.)

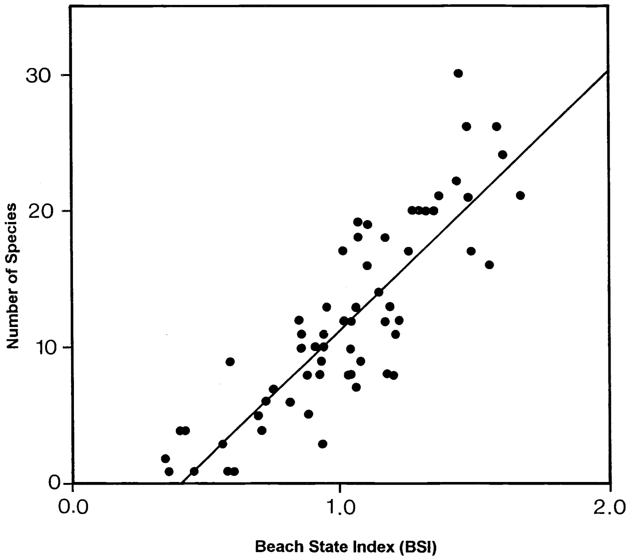


FIGURE 7 Compilation of species richness data from four continents, including wide-ranging surveys in Australia. (After Hacking, 1997.)

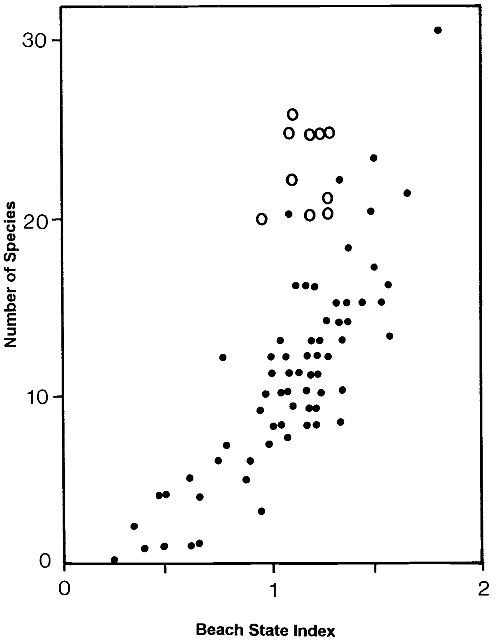


FIGURE 8 Data from tropical beaches in Oman (○) plotted against global data showing elevated species richness. (After McLachlan *et al.*, 1999.)

sample area per transect in the studies discussed here falls between 2 m^2 and 4.5 m^2, with most cases being around 3 m^2. The study by Jaramillo *et al.* (1995) concluded that a total sample area of 3 m^2 was sufficient to recover 90% of the species on microtidal beaches but would result in undersampling on wider beaches, such as macrotidal flats. This implies that actual richness on the latter beaches is higher than recorded in Figs. 5 and 6 and supports the suggestion of Hacking (1997) that the response of species richness to BSI may not be linear but may rather tend to be exponential over beaches with high BSI values. More work is needed to clarify this.

VI. LATITUDINAL AND OTHER EFFECTS

McLachlan (1990) suggested that, because wave energy is highest in temperate areas, these regions would enjoy a preponderance of dissipative beaches and hence greater species richness. Subsequently, Dexter (1992) proposed that tropical beaches harbor less diverse faunas than temperate beaches. However, this was based on empirical data and did not consider the role of morphodynamics. More recent work comparing beaches of the same morphodynamic state has indicated that tropical beaches may in fact support greater species richness than temperate beaches of the same type; exceptional species richness has been reported for beaches in Queensland and Oman (Fig. 8). Thus, in addition to the three physical factors listed earlier, latitude also appears to play a role in determining species richness. Certainly, tropical beaches do not seem to be poorer in species than temperate beaches of similar types.

A final factor of importance to intertidal beach macrofauna communities may be beach length. In Australia, it has been shown that angling activity and success, and thus intertidal bait organism communities, may become richer as beach length increases; a non-linear relationship between beach length and species richness has been found for Chilean beaches, and studies showing clam abundance increasing with beach length also support this. Experience suggests that short "pocket" beaches are often impoverished and this factor deserves further study.

VII. CAUSATIVE FACTORS

For ocean beaches, the number of species recorded in a single transect survey ranges from 1 to 40, if insects are excluded. Reflective beaches with coarse sand, dynamic swash action, and rapid drainage may be devoid of intertidal species and harbor only supralittoral forms, whereas more dissipative beaches and flats may harbor 20–40 species at least. Loss of species under reflective beach conditions occurs mainly in the lower intertidal, not the supralittoral zone. It is in the lower intertidal zone where most species are added as the shore flattens towards the dissipative or tidal flat extreme. So the question must be raised: Why are so many species excluded towards the reflective beach extreme?

Considering only between-beach patterns, that is, if whole beaches are compared, the physical environment impinging directly on beach fauna differs primarily in two sets of variables. These two variables are (1) the movement of water over and through the beach face, which has been termed the swash climate, and (2) the sediment particle size range. This ignores the normal intertidal gradient of exposure, which is relevant to any within-beach study.

On most beaches, waves do not break directly in the intertidal but rather in the surf zone. It is only after transformation to bores and crossing the surf zone that wave energy reaches the beach face as swash. Thus, intertidal fauna experience the effects of waves, tides, and the surf zone transformation of these forces as the swash "climate" on the beach face. McLachlan (1990) suggested that it was not the beach state or morphodynamic type itself that was important for the fauna, but rather the swash climate associated with it.

There is a consistent relationship between beach type and swash climate features. Dissipative beaches are characterized by swash with extended periods and lengths, variable speeds, most swash activity below the effluent line (water table outcrop), and fairly laminar swash flow. Reflective beaches display the opposite swash features: the swash climate is extremely harsh; there are high swash speeds throughout the tidal cycle; and waves break directly in the intertidal, resulting in considerable turbulence and increased probability of animals being stranded above the effluent line where unsaturated sand might make burrowing difficult. Swash drainage through the sand is also rapid on reflective beaches and there is little inundation time for feeding. Physical stress in the swash zone on the beach face thus increases from dissipative to reflective beaches.

The "swash control hypothesis" suggested that swash climate controlled beach macrofaunal community structure (McLachlan, 1990). McLachlan *et al.* (1993) refined these ideas to define the "swash exclusion hypothesis": the swash climate associated with dissipative beaches is considered sufficiently accommodating and varied to enable most psammophilic macrofauna species to maintain viable populations; however, as beach type changes through intermediate states toward reflective conditions, the increasingly inhospitable swash climate excludes more and more species until, in the fully reflective situation, only supralittoral forms (talitrid amphipods, ocypodid crabs, insects), which live "outside" the area affected by swash, remain. Swash also has important indirect effects on the macrofauna since animals move, feed, and reproduce in the water moving over the beach face. Swash patterns and percolation of water through the beach face above the effluent line are closely coupled.

The coupling between beach state and swash climate is clear, but there is no published study undertaking a simultaneous examination of swash climate and fauna. However, recent work in New Zealand shows a strong correlation between community parameters and swash climate. Further, it has been demonstrated that the hallmarks of bivalves that are able to survive on reflective beaches are small size and high density, both adaptations for stability under turbulent conditions. The harsh swash climate on reflective beaches thus appears to select for relatively small size and high body density in bivalves, but may select for larger size and mobility in crustaceans.

It is now clear that swash is not the only physical factor of importance to beach macrofauna; sand particle size also plays a role. Ongoing work shows that small body sizes tend to be excluded in coarse sediments and various authors have demonstrated how sand particle size influences burrowing rate in a variety of sandy beach species—in most cases coarse sand makes burrowing difficult or impossible. Sandy beach isopods, for example, have been shown to select finer substrates. Coarse sand can cause vicious abrasion, especially of bivalve shells, and it reduces sand saturation by raising permeability and hastening drainage.

In the only study to examine faunal community response to swash and sediment factors over a range of each types, Gary Stephenson and I have concluded that both sediment particle size and swash climate directly influence the number of species of macroinfauna on wave-exposed sandy beaches in northern New Zealand. Since both factors are closely related to beach type, this explains the good correlations between beach type and macroinfaunal species richness. In general, conditions of fine sediment and long swash periods, with limited swash activity above the effluent line, appear most conducive to developing rich faunas. The key characteristic of this type of swash may, in fact, be its low degree of turbulence and laminar flow over fine, saturated sand, enabling even delicate forms to survive. Thus Stephenson and I propose that the swash exclusion hypothesis should be modified to a hypothesis of physical exclusion of species by coarse sand, turbulent swash, and low effluent lines as one moves towards the reflective beach condition.

VIII. BIOLOGICAL FACTORS

There is little evidence of biological factors controlling intertidal beach macrofauna. Predation by birds, fishes, and invertebrates is well documented, indeed quantified (Brown and McLachlan, 1990), but it has not been demonstrated to influence species richness. The absence of an underlying base of competition probably precludes this. In such dynamic, three-dimensional environments, with highly mobile faunas, competition is limited and indicated in only a few exceptional cases.

One biological factor that can add to species richness is the input of seaweed wrack to a beach, if the input is not so great as to cause disturbance, smothering, and deoxygenation. In cases of moderate input, many wrack-associated species may be added to the usual beach fauna. A well-developed drift line can also add many insects. Progressing along the gradient from reflective, wave-dominated beaches to macrotidal flats where tidal factors are more important than waves, increasing stability of the substrate leads to the formation of permanent burrows and more complex communities with greater scope for biological interactions. At some point, as yet not defined, in the continuum between beaches and tidal flats, biological factors are likely to become more important than outlined here for wave-exposed beaches.

IX. DISCUSSION

Three paradigms provide perspective for our current understanding of the factors controlling the large-scale patterns of species richness of sandy beach intertidal benthic macrofauna.

The Autecological Hypothesis The autecological hypothesis of I. Noy-Meir states that inhabitants of physically controlled environments respond independently to the physical environment. This implies that the presence or absence of a species on a beach will be little influenced by biological factors.

Beach Morphodynamic Models Beach morphodynamic states can be defined by three physical factors: wave energy, tide range, and sand particles size. Increases in wave and tide energy and decreases in sand particle size all result in beaches becoming wider, flatter, and more dissipative.

Richness and Abundance Coupled to Beach State Intertidal sandy beach macrofauna increase linearly (probably) in species richness and exponentially in abundance in response to the increasing dissipativeness of beaches, that is, from microtidal reflective systems through high-energy dissipative systems to ultradissipative tidal flats.

Beyond these paradigms, two further hypotheses can be erected to explain the known patterns of sandy beach macrofauna biodiversity:

Latitude and Beach Length Tropical beaches support more species than do temperate beaches of the same type, and long beaches appear to support greater abundance and more species than do short (or pocket) beaches. The former trend may be due to the greater species pool available to colonize beaches in the tropics, whereas the latter trend reflects the more favorable surf circulation patterns on long beaches, which promote greater retention of particulate primary production, better recruitment of planktonic larvae, and larger and therefore more resilient adult populations. Long beaches are also more persistent, that is, unlikely to erode away completely during storms, as can happen on pocket beaches.

Swash and Sand Control Two physical factors define the immediate environment experienced by the beach macrofauna and may therefore control species richness: the swash climate and sand particle size. Reflective beaches have harsh swash climates in the sense of high turbulence, short swash periods, and rapid swash drainage, resulting in low effluent lines; beaches at the other extreme have more benign swash climates in the form of long-period swash/tidal bores with laminar flow over the beach face and low turbulence, keeping much of the beach saturated because of the slow drainage. All beach fauna can burrow into saturated sand and, in terms of grain size, fine to medium sands seem optimal for most beach fauna. Coarse sand appears to exclude small or delicate forms by crushing and abrasion and most species experience decreasing burrowing efficiency in coarse sands. Thus harsh swash climates and coarse sand associated with reflective beaches appear to exclude many species. The degree of exclusion of species by these factors will increase from the dissipative/macrotidal to the reflective/microtidal end of the continuum of beach morphodynamic types.

It has been stated that community structure in harsh environments is in nonequilibrium, characterized by the decoupling of loose biotic interactions and the independent response of species to environmental variations. Thus populations and communities in stressful environments may respond primarily to the abiotic environment, that is, be highly stochastic and thus loosely structured. This is a restatement of the idea of a physically controlled community and the autecological hypothesis and appears to be highly relevant to the case of sandy beaches. The dynamic, three-dimensional nature of wave-exposed ocean beaches, their shifting populations, and the absence of biological structures (they are even devoid of permanent burrows in most cases) leave limited scope for biological controls of community structure. Furthermore, the evidence presented here has shown that an index based only on the three physical variables that define a beach (tides, waves, and sand) has high predictive power to estimate the species richness likely to be encountered on ocean beaches. Latitude, beach length, and biological interactions may contribute additionally, to a small extent, to explain some of the variability in the data, but their roles have yet to be conclusively demonstrated in high-energy beach environments.

Several lines of research could improve our understanding of patterns of sandy beach biodiversity and their control.

1. The BSI is a simple index developed partly by ecologists and fitted to a few data sets. Collaborative work between ecologists and geomorphologists, using all the data available at present and investigating new regions, could lead to refinement of this to an even more appropriate index (or indices) with greater predictive power.
2. Comparative work contrasting similar beaches in the tropics and temperate latitudes is needed to confirm conclusively whether tropical beaches do indeed support higher diversity and what controls this.
3. Once these processes are better understood, useful predictions could be made of global beach diversity patterns based purely on the global distribution of tide and wave regimes and sedimentology. This capability would be useful in predicting and identifying areas of high diversity and in planning coastal conservation strategies.

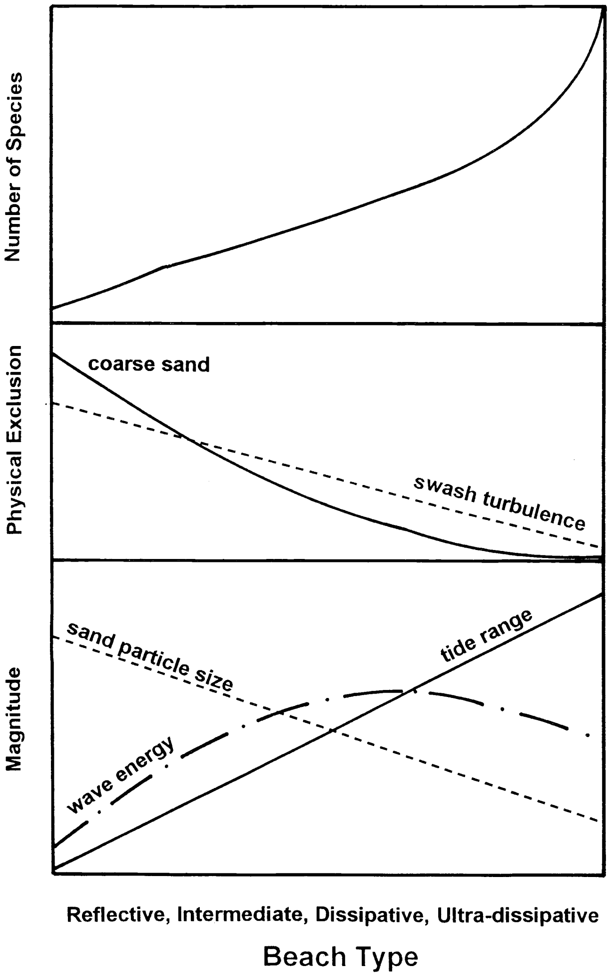


FIGURE 9 Simplified diagrammatic model of changes in defining variables (wave energy, tide range, and sand particle size), levels of physical stress induced by the substrate and swash, and species richness over a continuum from microtidal reflective to macrotidal ultra-dissipative sandy beaches.

X. CONCLUSIONS

In the case of wave-exposed, ocean sandy beaches, the species richness of the intertidal (and immediately supralittoral) macrobenthic fauna is strongly controlled by, and predictable on the basis of, the physical nature of individual beach environments rather than global biogeographic patterns. Beaches that tend toward the ultra dissipative/tidal flat extreme—the product of large tide ranges, vigorous wave action, and fine sand—harbor most sandy beach species likely to be encountered in a region. Progressing from this end of the beach continuum towards reflective conditions, that is, decreasing tide range or wave energy or increasing sand particle size, results in beaches becoming steeper and narrower and the increasing exclusion of species. On highly reflective beaches, intertidal macrofauna may be absent and only supralittoral forms remain. These relationships are illustrated in simplified form in Fig. 9.

The mechanisms by which species are increasingly excluded as one moves towards reflective conditions appear to relate primarily to two environmental variables, the swash climate and sand particle size. Decreasing swash period, increasing swash turbulence, lower effluent lines, and coarse sand exclude delicate, less robust, and slower-moving species. At present our knowledge of how biogeographic factors influence species richness on sandy beaches is limited, but it appears that tropical beaches may support more species than do temperate beaches of similar morphodynamic type. There is considerable scope for further research to clarify these patterns and to decipher cause and effect relationships.

See Also the Following Articles

ESTUARINE ECOSYSTEMS • INTERTIDAL ECOSYSTEMS • LAKE AND POND ECOSYSTEMS • MARINE ECOSYSTEMS • RIVER ECOSYSTEMS

Bibliography

Bascom, W. (1980). *Waves and Beaches,* updated ed. Anchor Press/ Doubleday, New York. Brown, A. C., and McLachlan, A. (1990). *Ecology of Sandy Shores*. Elsevier, Amsterdam.

Davies, J. (1972). *Geographical Variation in Coastal Environments*. Longman, London.

Dexter, D. M. (1992). The sandy beach fauna of Praia de Bruce, Bubaque, Guinea Bissau. *J. Biogeogr.* **19**, 59–66.

Hacking, N. (1997). Sandy beach macrofauna of eastern Australia: A geographical comparison. Ph.D. thesis, University of New England, Armidale, Australia.

Jaramillo, E., McLachlan, A., and Dugan, J. (1995). The species/area relationship on sandy beaches. *Mar. Ecol. Prog. Ser.* **119**, 311–314.

McLachlan, A. (1990). Dissipative beaches and macrofauna communities on exposed intertidal sands. *J. Coastal Res.* **6**, 57–71.

McLachlan, A., and Erasmus, T. (1983). *Sandy Beaches as Ecosystems*. Dr. W. Junk. The Hague, Netherlands.

McLachlan, A., and Jaramillo, E. (1995). Zonation on sandy beaches. *Oceanogr. Mar. Biol. Annu. Rev.* **33**, 305–335.

McLachlan, A., Wooldridge, T., and Dye, A. H. (1981). The ecology of sandy beaches in southern Africa. *S. Afr. J. Zool.* **16**, 219–231.

McLachlan, A., Jaramillo, E., Donn, T. E., and Wessels, F. (1993). Sandy beach macrofauna communities and their control by the physical environment: A geographical comparison. *J. Coastal Res.*, Special Issue, **15**, 27–38.

McLachlan, A., De Ruyck, A. M. C., and Hacking, N. (1996). Community structure on sandy beaches: Patterns of richness and zonation in relation to tide range and latitude. *Rev. Chilena Hist. Nat.* **69**, 451–467.

McLachlan, A., Fisher, M., Al-Habsi, H. N., Al-Shukairi, S. S., and Al-Habsi, A. M. (1998). Ecology of sandy beaches in Oman. *J. Coastal Cons.* **4**, 181–190.

Peterson, C. H. (1979). Predation, competitive exclusion, and diversity in the soft sediment benthic communities of estuaries and lagoons. In *Ecological Processes in Coastal and Marine Ecosystems* (R. J. Livingston, ed.), pp. 233–264. Plenum New York.

Reise, K. (1985). *Tidal Flat Ecology*. Springer-Verlag, Berlin.

Short, A. D. (1996). The role of wave height, period, slope, tide range and embaymentisation in beach classification: A review. *Rev. Chilena. Hist. Nat.* **69**, 589–604.

COEVOLUTION

Douglas J. Futuyma and André Levy
State University of New York at Stony Brook

GLOSSARY

character displacement Pattern of geographic variation in which a character differs more between sympatric than between allopatric populations of two species.

competition Interaction between individuals of the same or different species, whereby resources (e.g., food, space, mating partners) used by one are made unavailable to others.

mutualism Symbiotic relation in which each of two species benefits from the interaction.

resistance An individual's capacity to reduce the damage inflicted upon it by an enemy.

specialization Evolutionary adaptation in a particular mode of life or habitat.

tolerance An individual's capacity to sustain damage by an enemy with limited decrease in fitness.

trade-off Negative correlation between traits, such that a benefit due to changes in the value of one trait is associated with a cost produced by changes in the value in another trait.

virulence Degree to which a parasite reduces the probability of survival or the reproductive capacity of an individual host. A relatively avirulent (benign) parasite has little impact on its host's fitness.

ALTHOUGH THE TERM COEVOLUTION is sometimes used to describe the joint evolution of different genes or characteristics within a species, it usually refers to *evolutionary changes in two or more different species owing to ecological interactions between them*. These interactions include competition, mutualism, and "enemy/victim" interactions between predators and prey, parasites and hosts, and herbivores and plants. Even in this restricted sense—the subject of this article—coevolution has several meanings.

I. MEANING AND VARIETIES OF COEVOLUTION

(1) *Patterns of correlated evolution*, especially *phylogenetic congruence*, among species or other lineages are usually detected by phylogenetic analysis. In extreme cases, the phylogeny of one clade of species, such as

host-specific parasites, is congruent with that of a clade with which it interacts such as host species (Fig. 1a). Such congruence would require *coincident speciation* of hosts and associated parasites. Deviations from perfect congruence might be caused by speciation of a lineage but not its associate, by extinction of one but not the other, or by a parasite switching from one host to another (colonization) (Fig. 1b).

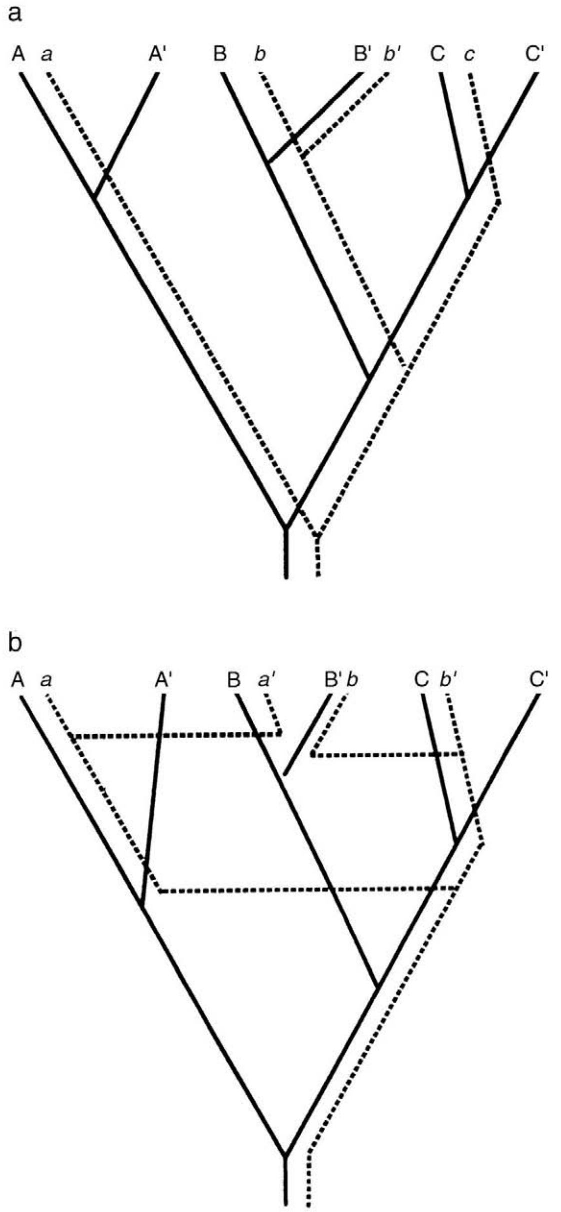


FIGURE 1 Congruent (a) and incongruent (b) phylogenies of hosts (solid lines) and parasites (dashed lines). Letters indicate most closely related taxa. Each parasite taxa is specialized on the host with which it is closely associated in the diagram.

In some cases, the coincident speciation that underlies concordant phylogenies may be caused by the interaction between associated species. However, concordance might stem solely from a shared history of geographic isolation, without the interaction having played any causal role. Although phylogenetic concordance does not in itself provide evidence of reciprocal adaptation of the interacting species, it does imply extended historical opportunity for adaptation to have occurred. The greatest phylogenetic concordance has been described for some maternally transmitted endosymbiotic bacteria and their eukaryote hosts, and for certain ectoparasites (e.g., lice of gophers) that are primarily transmitted by contact among conspecific hosts. In many groups of parasites and herbivorous insects, related species are associated with related hosts, that is, members of a clade, but the phylogenies of the two clades are not congruent. This pattern often implies a long history of association, but one that has included shifts of parasite lineages among host lineages.

(2) Coevolution consisting of *reciprocal adaptive responses* has been rather arbitrarily classified as *specific* (or pairwise, or tight) coevolution and *diffuse* (or guild) coevolution. Specific coevolution is reciprocal adaptation of two species, independent of their interactions with other species, whereas diffuse coevolution occurs when evolutionary change in one species affects its interaction with two or more other species (e.g., a change in a host that affects two parasite species). Specific coevolution is easier to visualize, model, and detect. Diffuse coevolution is most likely to occur when a species' response to selection by one interacting species is accompanied by genetically correlated effects on its interactions with other species. The term "diffuse coevolution" has also been used to describe nonadditive selection on one species by two or more other species, owing to an interaction between them or between their effects on the recipient species.

(3) *Cospeciation* may be induced by the interaction between species, but this probably requires special conditions. A likely case is in figs (*Ficus*) and the highly host-specific seed-eating wasps that pollinate them. Divergence between population of either a fig species or a wasp species in features governing the host preference of the wasp may engender divergence in the fig, resulting in reproductively isolated populations of both. Such cospeciation may result in concordant phylogenies.

(4) *Escape-and-radiate coevolution* describes a scenario in which an evolutionary lineage diversifies after it evolves a defense that breaks its association with enemies. Such a scenario was first proposed by P. R.

Ehrlich and P. H. Raven for plants and herbivorous insects. In this hypothesis, a plant lineage that evolves a novel defense sheds most of its herbivores, and therefore (by an unspecified causal inference) diversifies into many species. Only later do some herbivore species adapt to one or more of the plant species, and diversify in turn.

Collectively, these hypothetical scenarios suggest that coevolution may increase both the numbers of species and the phenotypic diversity (disparity) of species by selecting for the many characteristics that affect their interactions. Coevolution may therefore be an important engine of biological diversity, but we shall shortly note that coevolution may also cause extinction.

Often, the adaptation of one species to others is obvious, but reciprocal adaptation of the others is not. Because we seldom know a priori whether or not adaptation has been reciprocal, adaptations to ecological interactions are often loosely referred to as "coevolutionary," even if evidence for reciprocity is slight or lacking. Certainly the *study* of coevolution includes adaptations to other species, which may or may not prove to have adapted in turn.

II. EVIDENCE OF COEVOLUTION

A. Sources of Evidence for Coevolution

The study of coevolution comprises the same approaches as studying evolution in general. As in the broader field, the first forms of evidence for coevolution consisted of detailed natural history observations, descriptions of the diversity of adaptive structures that mediate ecological interactions, and comparison among populations and species.

Charles Darwin published the first comprehensive illustration of how intricately species are adapted to one another and how structural complexity can be explained by coevolution in his description *The Various Contrivances By Which Orchids Are Fertilized By Insects* (1877). By comparing the shape of different orchid flowers and their associated pollinators, Darwin demonstrated that certain moth features are needed to successfully obtain nectar from the flower, features that are indeed exhibited by their specialized pollinators. By restricting nectar collection to a few pollinators, orchids increase the likelihood of cross-fertilization. Such is the correspondence between flower and pollinator shape that on observing the 29-cm-long nectar-bearing spur of the Madagascan orchid *Angraecum sesquipedale,* Darwin predicted the existence of a pollinating moth with a proboscis of that length. Such a moth, *Xanthopan morgani praedicta,* was indeed discovered 40 years later.

The description of patterns in plant use by lepidopteran larvae preceded the concept of *escape-and-radiate coevolution*. Higher taxa of butterflies often feed upon a single group of flowering plants. While some feed on more than one plant family, these tend to be closely related or have similar biochemistries. For instance, larvae of the butterfly subfamily Pierinae, or whites, feed predominantly on the families Capparaceae and Brassicaceae, which are closely related. Some whites also feed on members of the family Tropaeolaceae that share with the other families the production of mustard oil glycosides and a rare fatty acid. These regularities imply an important role for plant secondary metabolites in determining butterfly host use. Given that these compounds affect herbivore behavior, acting often as deterrents, secondary chemistry may have constituted the key feature that allowed plant escape.

Comparisons among conspecific populations have also been suggestive of coevolution. The coloration pattern of the butterfly *Heliconius erato*, thought to be a signal to predators indicating distastefulness, varies among populations in Central and South America. Strikingly, the wing coloration of *H. melomene*, an equally distasteful congener with distinct life history and host preference, varies geographically in parallel with *H. erato*. This pattern is thought to be an example of coevolution of mimicry between prey species that share a predator. Fritz Müller, a contemporary of Darwin, first suggested this particular model of coevolution to explain similarities in wing pattern among unpalatable butterfly species belonging to two distinct genera (*Ituna* and *Thyridia*).

Müller also introduced the use of mathematical models to study the coevolutionary process. Modern mathematical and computer simulation models may incorporate population genetics, quantitative genetics, evolutionary game theory, and optimality theory. Mathematical modeling has proven useful in describing the dynamic of the interactions betwen species and in determining which conditions favor coevolution.

Although ecological interactions usually do not "fossilize," the analysis of paleontological records has provided some evidence of coevolution. For instance, the appearance in the Ordovician of predaceous cephalopods is associated with the simultaneous appearance of several defensive strategies on the part of their prey (e.g., strong sculpture and coiling in gastropods and shell-bearing cephalopods, spines in echinoderms), suggestive of diffuse coevolution between predators and their prey. The antiquity of certain interactions may also be determined by inspecting fossils of extant species. Several plant families possess structures (domatia) that harbor mites, which attack plant enemies. Domatia similar to the modern form have been discovered in fossilized leaves from the Eocene, 55 million years ago!

The relative age of clades of associated taxa is relevant for demonstrating correlated coevolution or cospeciation. These processes would be necessarily excluded if one group were much older than the other. The age of an association, or of interaction-related adaptations, can often be estimated from phylogenies with time calibration (e.g., using approximate molecular clocks or stem-group fossils). Molecular evidence from deep-sea vesicomyid clams and the sulfur-oxidizing endosymbiotic bacteria on which they depend for nourishment indicates that the interacting clades are both approximately 100 million years old. These two lineages appear to have been in close association since their origin and to have cospeciated, as indicated by the remarkable level of congruence between their estimated phylogenies. Phylogenetic information also becomes relevant in testing whether a character is an adaptation for an ecological interaction or an ancestral feature that exists in the absence of the interaction.

In some instances it has been possible to document the particular genes that affect a species' interaction. H.H. Flor found several genes in flax (*Linum usitatissimum*) that provide resistance to the rust *Melampsora lini*. Rust virulence is determined by a set of complementary genes, in a one-to-one relationship. This study inspired the gene-for-gene model (see Section VI,A), which has become a paradigm of phytopathology. Most traits, however, have a complex genetic basis, involving many genes. Such complexity requires a quantitative genetic approach, which partitions the trait variation into genetic and environmental components. This approach has demonstrated that many of the traits relevant to interactions have genetic variability, that is, there is potential for coevolution. For example, the wild parsnip (*Pastinaca sativa*) and its most important associated herbivore, the parsnip webworm (*Depressaria pastinacella*), are thought to be engaged in coevolution mediated by the evolution of furanocoumarins and the insect's detoxifying mechanisms. May R. Berenbaum has documented genetic variation both in the production of furanocoumarins and in the webworm's ability to metabolize this group of plant toxins.

Quantitative genetics is also used to measure correlations between traits. The detection of negative genetic correlations is indicative of trade-offs between traits, such that selection for the increase in value of one trait leads to a decrease in value of the correlated trait. Tradeoffs are of particular relevance in explaining evolutionary constraints, and particularly why species are specialized. Pea aphid (*Acyrthosiphon pisum*) clones collected from two crop plants (alfalfa and red clover) exhibited higher fitness when reared on the plant from which they had been collected, suggesting local adaptation. The negative genetic correlation in fitness across crops may constrain the evolution of generalist clones, as these would be outcompeted on either plant by crop-specialized clones.

Measurement of genetic variation and correlation between traits offers information on the genetic context in which selection can act. Correlations between traits and fitness suggest the form and direction of selection. In a greenhouse study, the wild parsnip exhibited a negative genetic correlation between concentration of several furanocoumarins and seed set, suggesting that the production of the chemicals may impose a cost to reproduction in the absence of the parsnip webworm. These negative correlations were not detected in the field, indicating that presence of furanocoumarins increases fitness in the presence of the herbivore. Ideally such studies are performed in a natural setting, as our ultimate interest is understanding how natural selection works in the wild, but one can use model systems in the laboratory, such as evolving populations of bacteria and bacteriophage.

Finally, studies of interacting species are commonly based on the analysis of single communities. However, most species are composed of many local populations, and increasing importance has been attributed to the geographical structure of species and their interactions. Across the distribution of an interaction one is likely to observe a mosaic of selection pressures as a result of variation in abiotic and biotic factors, and the particular demographic and genetic histories of local populations. Some localities may be coevolutionary hot spots, that is, sites of reciprocal coevolution, whereas in others selection may be unidirectional or act on neither species. The geographic variation in outcomes is further modified by gene flow among populations. Consequently, different degrees of coadaptation are to be expected among populations. Clearly our understanding of the dynamics of a species interaction requires the study of many communities and interpopulation processes.

A few cases that satisfy the requirements of long-term multipopulation studies have emerged recently and have reinforced our need for a *geographic mosaic theory of coevolution*. For instance, resistance and virulence structures of *Linum* and *Melampsora* (referred to earlier), studied in New South Wales, sometimes vary dramatically across populations and time. The frequency of susceptible genotypes of flax will affect the local frequency of a particular strain of flax rust, but additional factors were found to play a role, namely, drift, extinction, and migration from neighboring populations. The geographic structure of flax and rust proved to be an essential factor in explaining the persistence of the interaction.

B. Rates of Adaptation

Phylogenetic reconstructions and the fossil record provide some examples of ancient interactions between groups of organisms. Character mapping of feeding habits onto beetle phylogeny indicates that conifers and cycads are the ancestral host plants. The oldest fossils of herbivorous beetle taxa, from the Jurassic, are indeed ancestral forms of modern conifer and cycad feeders, suggesting that this association may have evolved 200 million years ago, making it the oldest insect/plant association known. Lichens, which are an intimate association between fungi and green algae, may be 100–200 million years old. Yet how often do new interactions establish themselves, how rapidly do species adapt to one another, and does that process occur on a timescale that is relevant to ecological processes?

The evolution of interspecific interactions can occur very rapidly, that is, during the course of 100 years. Examples include cases where recorded changes in the environment led to measurable changes in phenotype. In 1977, the Galápagos Islands suffered a severe drought during an El Niño event, causing an increase in the median size of available seeds. During the same event, the medium ground finch (*Geospiza fortis*) suffered strong selection for an increase in bill size, as only larger-billed birds could effectively feed on the available large seeds. This morphological shift occurred in a single generation. Many cases involve species of animals and plants introduced into new habitats. Populations of the apple maggot (*Rhagoletis pomonella*) that fed exclusively on native hawthorn have in the course of the past 100 years become adapted to a novel host, the introduced apple. The loss of interaction-related adaptations has also been observed to occur on a short timescale. Within the course of a 34-year study, guppies (*Poecilia reticulata*) artificially introduced into drainages in Trinidad, where natural predators are absent, lost defensive antipredator behaviors exhibited in the source populations.

Although these examples unequivocally demonstrate that interspecific interactions can evolve rapidly, they do not provide examples of coevolution. Almost all examples describe the evolution of *one* species adapting to another, with little reference to reciprocal changes. One of the few exceptions is the effort to biologically control the rabbit (*Oryctolagus cuniculus*) in Australia. The introduction of the myxoma virus into Australia initially reduced the rabbit population drastically. Such strong selection led to the evolution in just a few generations of increased resistance in rabbits. This evolutionary change was accompanied by a decline in the virulence of the virus to an intermediate level. Although in this case evolutionary change was observed in both interacting species, this does not necessarily constitute a case of coevolution. The reduction in virulence did not occur as a result of changes in rabbit resistance, but rather because the initial strains of the virus were so virulent that they quickly killed the host, lowering the likelihood of transmission by the disease's vector. More recent data, however, seem to indicate that the initially more virulent strains are presently increasing in frequency, perhaps as a result of increased rabbit resistance.

Rapid evolution is not, however, unconstrained. Each species is influenced by its phylogenetic history, and even populations of the same species will possess different evolutionary potential. To grasp how often new interactions are formed, one needs to understand these constraints. Rapidly evolving interspecific interactions can potentially affect the dynamics of community structure. Models of community ecology should not rest on the assumption that evolutionary dynamics occur on a larger timescale that is irrelevant to ecological processes. But it has yet to be determined under which conditions coevolution drives community dynamics and whether it has a stabilizing role.

III. COMPETITION BETWEEN SPECIES

Hypotheses about the coevolution of competing species have long played a major role in both evolutionary biology and ecology. Most of the hypotheses are variations on the theme that species should evolve differences in resource utilization, thus reducing competition between them. For Darwin, the principle of "divergence of character" in response to competition, described in *The Origin of Species*, explained why an ancestral species should give rise to multiple, diverse descendant species. W. L. Brown and E. O. Wilson cited reduction of competition as a major cause of *character displacement*, the term they coined for a greater difference between sympatric than allopatric populations of two species. D.

Lack and R. H. MacArthur provided famous examples of differences in diet and foraging behavior or trophic morphology among closely related species of birds, interpreting the patterns as stemming from avoidance of competition. G. E. Hutchinson cited instances of apparently constant ratios of the dimensions of trophic structures among sympatric congeners. MacArthur and R. Levins elaborated on this theme, developing a theory of limiting similarity that concerned the expected degree of niche differentiation among sets of potentially competing species, as part of a theory of the evolution of community structure that has had lasting influence. The importance of competition and of coevolution of competitors was the subject of vigorous debate in the 1970s and 1980s.

The most detailed recent models of coevolution of competitors (Taper and Case, 1992) assume that the size of a phenotypic feature, such as body size or beak depth, is correlated with the mean food type (e.g., prey size) that an individual uses. The evolution of the character, and thus of resource use, is modeled using quantitative genetic models for polygenic traits. (The character is approximately normally distributed, with specified additive genetic variance and heritability.) It is assumed that "a jack of all trades is master of none," that is, that there exist trade-offs in adaptation so that each phenotype performs effectively over only a relatively small range of resources (its within-phenotype niche width, WPNW), and that a specialized phenotype (with smaller WPNW) is more efficient than a generalist. (For example, studies of finches have shown that the rate of energy intake from smaller versus larger seeds is greater for small-billed versus large-billed birds, respectively.) In the simpler models, the competitive effects of phenotypes on each other are symmetrical and are proportional to the degree of resource overlap, whether the phenotypes be the same or different species. Thus competition is more intense when two phenotypes are more similar.

The size–frequency distribution of resources is assumed to be unimodal. Thus the rarity of resources in the tails of the distribution selects against extreme phenotypes. Consequently, the mean phenotype of a solitary species evolves to match the peak of the resource curve. Frequency-dependent selection maintains variant phenotypes, each adapted to a slightly different resource. However, the higher WPNW is, the less genetic variation is maintained. Because of recombination among the many loci affecting the trait, the frequency distribution of phenotypes remains approximately normal. It does not necessarily fit the frequency distribution of resources, some of which may be underutilized.

When applied to two or more species, these models generally support earlier suppositions. In each species, phenotypes that do not bear the combined burden of interspecific and intraspecific competition have higher fitness. Hence, the species evolve so that their phenotypic means are spaced apart (character displacement). Moreover, the phenotypic variance within each species at equilibrium is somewhat smaller than in a solitary species. At equilibrium, some overlap remains in the species' resource utilization, for otherwise abundant food items of intermediate size would be left unused, creating selection for genotypes that can use them. Under some conditions, especially if two species initially overlap greatly in resource use, they may converge. Competitive exclusion, resulting in extinction, then becomes more likely. Moreover, if competitive effects are asymmetrical (e.g., if larger individuals reduce the fitness of smaller ones more than the converse), a species may converge toward another and "chase" it to extinction.

Facile invocation of evolutionary responses to competition to explain ecological patterns was severely criticized in the 1980s, resulting in more critical evaluation of evidence. Even under the new scrutiny, many data strongly support the theory of coevolution of competitors. For example, character displacement has been documented in many instances, such as sticklebacks in British Columbian postglacial lakes. Some lakes harbor two species, recently derived from a common ancestor, that differ in microhabitat and in morphological features associated with feeding. Other lakes have only one species, with intermediate behavior and morphology. Closely related sympatric species often differ more consistently than would be expected by chance. For instance, body size in bird-eating hawks (*Accipiter*) is correlated with average prey size, and throughout the world sympatric species of these hawks differ more in size than do random pairs of *Accipiter* species.

Such nonrandomness of sympatric assemblages might arise from either coevolution or ecological assembly (colonization followed by extinction of excessively similar species). Phylogenetic analyses of anoline lizards in the West Indies have provided evidence of both processes. Many islands in the Lesser Antilles have both a large and a small species. These form distinct clades, so there is no evidence that the size of any one of these species have evolved in response to a sympatric congener. But each of the Greater Antilles has a monophyletic group of species that differ ecologically and morphologically from each other in a parallel pattern that is almost the same on each island. For example, each island has at least one species of "trunk anole," one “twig anole,” and one “crown giant.” Coevolution has taken much the same historical course in each case.

Release from competition is thought to have important evolutionary consequences. “Ecological release” at a microevolutionary level is illustrated by cases in which only a single member of a genus occurs in a region, and there occupies a broader ecological niche than do species that coexist with congeners elsewhere. The single species of finch on isolated Cocos Island, northeast of the Galápagos archipelago, feeds on a much wider variety of items than does any of its many Galápagos relatives, and the sexual dimorphism in beak size and forging mode is greater in the Hispaniolan woodpecker than in its continental relatives that are sympatric with other woodpecker species. At a macroevolutionary level, the spectacular diversification of African lake cichlids in the virtual absence of other fishes, the explosive diversification of mammals after the K-T mass extinction of large reptiles, the flowering of modern turtles after the extinction of primitive amphichelydian turtles, and many other examples are thought to show how diversification can be released when incumbent competition is alleviated.

IV. RESPONSES TO COMMON PREDATORS

A. Escape Space and Divergent Defenses

Competition is not the only possible explanation of resource partitioning and community assembly. For instance, species may interact indirectly via a shared predator. Predator density is highest where prey species coexist, and consequently a particular prey species will be maintained at lower density if it coexists with other prey species than if it is the sole victim. This form of predator-mediated *apparent competition* creates a pattern identical to that of direct competition. Coexistence is possible if prey species partition the resource they compete for, in this case an ecological space of reduced predation, known as an *escape or enemy-free space*.

Predators tend to increase their search intensity with increasing densities of prey. Search behavior can be improved by specialization in a foraging strategy, for instance, by attacking prey of a certain shape (i.e., similar search image) or prey that occupy particular sites, regardless of whether they belong to the same or different species. If faced with a diversity of prey patterns, they are most likely to form a search image for the most common pattern. This has been demonstrated for many bird and mammalian predators that use primarily visual cues. For example, if wild thrushes are presented with different combinations of two morphs of the snail *Cepaea nemoralis,* they will invariably preferentially attack the most common morph. The rarer individuals that differ from the ecological characteristics that the predator is attuned to will consequently experience lower predation rates. Such frequency-dependent selection within species, or *apostatic selection,* could lead to the divergence among prey species and an eventual increase in the diversity of forms and coloration patterns, that is, *aspect diversity*.

Evidence for this process is mostly circumstantial R. E. Ricklefs and K. O’Rourke compared morphology, color, and behavioral characteristics of the moth species of a tropical and two temperate communities. Despite a higher number of moth species in the tropical site, the average similarity among moths within a community was the same in all three sites. They suggested that predation influences competition among moths for “escape space,” limiting the similarity of appearance in different moth communities. *Cepaea nemoralis* and *C. hortensis* are both polymorphic for shell coloration pattern, and they share similar patterns. Bryan Clarke studied mixed colonies of *Cepaea* and found a negative correlation between the frequencies of visually similar shell patterns in the two species.

B. Mimicry

Another form of coevolution among prey species subject to common predators is the coevolution of mimicry. Two common forms of mimicry are typically distinguished. Batesian mimicry refers to the convergence of palatable mimic species on distasteful models. Predators learn to avoid certain prey shape and color patterns they experienced as distasteful and mimics of such patterns can profit from this aversion. Monarch butterfly larvae, *Danaus plexippus,* feed almost exclusively on milkweed, from which they sequester cardiac glycosides. These toxic compounds are retained in the adult and vertebrate predators quickly learn to avoid both monarch adults and the more palatable mimetic viceroy butterfly, *Limenitis archippus*. Cleaner fish provide a variation on Batesian mimicry. In coral reefs in the Pacific, many fish allow cleaner fish, such as the sea swallow (*Labroides dimidiatus*), to feed on parasites on their bodies and even in the interior of their mouths. The sabre-toothed blenny (*Aspidontus taeniatus*) mimics the white-and-black-striped coloration and swimming pattern of *Labroides*. By taking advantage of the passive behavior of fish toward the model, it is able to approach fish and bite off pieces of tissue. *Labroides* and *Aspidon-*

tus show parallel variation in color patterns across different geographic areas, which strongly suggests that indeed the mimic is converging on the model. It is a matter of discussion, however, whether mimics will instill evolution in the model, which might be expected to evolve differences that lessen the resemblance.

Müllerian mimicry refers to the convergence toward a similar pattern among unpalatable species. Faced with several undesirable species that look alike, a predator must learn a lower number of patterns to avoid. Evolution in all prey species leads toward a common pattern, and so warrants the designation of coevolution. One of the most striking cases of Müllerian mimicry, mentioned earlier, is the convergence between the neotropical butterflies *Heliconius erato* and *H. melpomene*. Despite differences in life history, these species share a common wing color pattern that varies geographically in parallel. One of the species, *H. erato,* is usually the most abundant where both species co-occur, raising the possibility that parallel evolution occurred by mere convergence of the rarer *H. melpomene* toward a common model. However, comparison between sympatric and allopatric populations of *H. erato* in Central American revealed that the width of the *H. erato* yellow hindwing bar converges upon that of *H. melpomene* when in sympatry, suggesting that both species converge on each other.

V. ENEMY–VICTIM INTERACTIONS

The general heading of enemy–victim interaction includes a variety of antagonistic interactions, ranging from those in which the enemy (or predator) usually kills several individuals (prey) to sustain itself, to scenarios in which the enemy (or parasite) restricts its negative effects to one victim (host) during most of its lifetime. The interaction between plants and herbivores is multiform, depending strongly on the herbivore taxon, and does not conform neatly to the previous division. Herbivores can be more akin to predators or parasites depending on whether they consume several individuals, as do seed harvesters, or complete their life cycle on a single plant, as do many leaf-feeding insects. They may also be quite unlike either category, as are many grazers and browsers that feed on parts of many individual plants without necessarily killing any of them. Differences in the degree of specialization of the enemy and victim may enhance the asymmetry of the interaction. Most plants, for instance, are attacked by numerous herbivores and parasites, yet many of their enemies are relatively specialized on particular plant taxa. These distinctions and characterizations are relevant as the nature of the interaction determines the selection pressures imposed on both species, and the resulting coevolutionary process.

Enemy–victim coevolution has been envisioned as an arms race in which the exploiter evolves offensives that increase the strength (i.e., frequency, intensity) of the interaction, and the victim evolves counterdefenses to decrease it. In predator–prey interactions, for instance, one would focus on traits that in some way affect the predator's functional response, that is, its rate of predation. This includes everything from the predator's detection ability and pursuit speed to the prey's aposematic or cryptic coloration and escape speed. An *evolutionary arms race* would correspond, for instance, to a continuous improvement of a prey's escape speed and the predator's pursuit speed.

Clearly there are limits to such continuous escalation. There are physical limitations to improving an organism's features, for example, to how fast an animal can run. Furthermore, investment in an interaction-related trait may imply costs. Development of the trait may require resources that could otherwise be invested in reproduction and other functions. For instance, elaborate morphological defenses (e.g., a thick shell or thorns) may imply an *allocation cost* expressed as a slower growth rate. Chemical defenses can act as feeding deterrents, but at high concentrations they may be autotoxic, implying a *physiological cost*. Additionally, some chemicals that act as deterrents toward generalists attract specialist enemies. *Ecological costs* produced by such genetic correlations limit a victim's potential to adapt to its complete array of enemies. Traits tend to affect more than one function of the organism, and a compromise must be reached between its role in the different functions. For instance, Geerat Vermeij suggests that such an adaptational dilemma could have been responsible for the extinction of most ammonite cephalopods. The external shell of cephalopods, other than providing passive protection, also affects the speed of locomotion and the ability to compensate for changes in pressure during vertical movements. Toward the end of the Cretaceous, as predation and competition became more intense, no further improvement in these functions could be reconciled by evolution of the shell.

Theoretical work has shown that the existence of costs and density dependence (the effect of a species' density on fitness of an individual of that species or of an interacting species) can hamper the occurrence of continuous reciprocal responses implicit in an arms race. Neither the prey nor the predator population is expected to always increase its investment in predation-related traits as a response to an increased investment by its partner. On the one hand, depending on the shape of the predator's cost–benefit function, the benefits reaped by an individual predator by investing more in predation can be insufficient to compensate for any additional costs. The predator's capture rate may remain unaltered if the evolution of greater prey elusiveness is accompanied by an increase in prey density. On the other hand, models have shown that the evolution of greater prey-capture abilities can lead to a decrease in the predator's equilibrium population size as a result of overexploitation of prey (Abrams, 1990). A smaller predator population can imply reduced selection for antipredator traits, and if the costs of defense become a more important selective factor than predation, the prey population may evolve a lower level of defense. So rather than a protracted arms race, one might expect enemy and victim to coevolve toward an intermediate stable state.

The models referred to previously focused on character values and population dynamic parameters. A different class of models focuses on the genetics that underlie interaction-related characters, while usually ignoring changes in population densities. For example, gene-for-gene models of host–parasite interactions examine changes in genotypic frequencies at a virulence locus in the parasite and at a resistance locus in the host. Coevolutionary cycles are expected under a broad range of conditions. As the frequency of a virulent genotype increases, selection for resistant host genotypes becomes more intense. Frequency-dependent selection will generate fluctuations in genotypic frequencies, and both populations will remain polymorphic. Some polygenic models also produce stability, as either stable limit cycles or equilibrium points, particularly if the additive genetic variances of the characters are high in both interacting populations. Stability is most likely if the victim has a larger genetic variance and is under stronger stabilizing selection than the enemy is. A higher additive genetic variance allows the prey to respond to selection more rapidly than the predator.

The "Red Queen" model extends the notion of evolutionary arms race to the level of a community. According to these models, evolution of co-occurring species can lead to the continuous deterioration of a species' environment, forcing it to constantly evolve just to avoid extinction. The name owes its origin to the character in Lewis Carroll's *Through the Looking Glass* who explained to Alice that one must run as fast as possible just to stay in place. Analogously, a community at evolutionary equilibrium determined by the Red Queen principle contains a set of interacting species continually coevolving at rates that exactly balance each other. It has been hypothesized that Red Queen coevolution has selected for recombination and sexual reproduction, as only those populations with available genetic variation could have sustained continuous coevolution for extended periods.

The most plausible models of enemy–victim coevolution predict either evolution to an equilibrium or an oscillatory coevolutionary "chase" (rather than indefinite escalation of offensive and defensive properties). The clearest evidence for these models' predictions should come from direct observation of long-term coevolution, which to date has been possible only with laboratory cultures of organisms with very short generations, such as bacteria and virulent bacteriophage. The most extensive coevolution observed in such experiments consisted of fixation of a resistance mutation in the bacterial population, then of a countervailing "virulence" mutation in the phage, and finally of a second resistance mutation in the bacteria, resulting in stable coexistence and an apparent genetic equilibrium. The resistance mutations of the bacteria carried costs that reduced fitness in the absence of phage, but evolution at modifier loci later reduced the cost. The failure of the phage population to coevolve greater virulence was attributed to architectural constraints on these very simple organisms.

Although data from natural populations reveal little about the long-term dynamics of coevolution, they do show that it occurs. Perhaps the most famous instance is the evolution of greater resistance of rabbits to myxoma virus, released as a rabbit-control measure in Australia. This was coupled with evolution of somewhat lower virulence in the virus (see earlier discussion). Comparisons among geographic populations and species also provide evidence of coevolution. Parasites are commonly better adapted to their local host population than to other populations, as in the case of a microsporidian that infects the cladoceran *Daphnia magna*. Brood-parasitic cuckoos are polymorphic for egg color: females generally specialize on one or another species of host, and lay eggs that mimic those of the favored host. Some species of birds, especially those most frequently parasitized, reject foreign eggs if they can detect them. In some such species, rejection behavior is stronger in populations that experience parasitism by cuckoos than in those that do not. Thus, both the cuckoo and some of its hosts have evolved responses to the interaction.

Geographic populations of interacting species sometimes vary in parallel, although the causes of variation can be difficult to identify. For example, the parsnip webworm (*Depressaria pastinacella*), the sole specialized herbivore that has accompanied wild parsnip (*Pastinaca sativa*) from Europe to North America, exerts selection on the plant's profile of several toxic furanocoumarins. Variation in furanocoumarin profile among populations of the plant is paralleled by variation in the differential capacity to metabolize these compounds in the associated populations of webworm. Although geographic variation in the insect is surely an adaptation to local plant characters, it is harder to show that variation among plant populations has evolved in response to selection by the insect.

Examples of adaptations of predators to prey or of the reverse abound. Yet in few instances has it been demonstrated that these adaptations have evolved in response to any one species. Many characteristics have a similar effect on several or many species of antagonists, and probably represent diffuse coevolution. For instance, the "Mesozoic marine revolution," in which the evolution of diverse shell-crushing crustaceans and fishes was mirrored by the evolution of spines, thicker shells, and other defensive features in many lineages of molluscs, doubtless represents diffuse coevolution. It is difficult in such cases to show that an evolutionary change in any one species stimulated coevolutionary change in another.

VI. SPECIAL FEATURES IN PARASITE–HOST COEVOLUTION

Several topics loom large in the study of parasites and hosts that seem to play a lesser role in coevolution of animal predators and their prey.

A. Gene-for-Gene Systems

Whereas features that mediate predator–prey interactions, such as size, trophic structures, and cryptic coloration, are usually polygenic (quantitative) characters, specific loci for resistance have been identified in several species of plants. For each such resistance gene, a corresponding locus in a fungal pathogen confers "virulence," meaning here the ability to attack and develop on a host with a specific resistance allele. (Most well-studied gene-for-gene systems involve crop plants, which has led some authors to suggest that resistance genes with large effects may be a result of the methods used to breed for resistance.) In the absence of pleiotropic costs, resistance alleles at all loci would be fixed in the host population and virulence alleles would be fixed in the parasite. If these alleles do have costs, then frequency-dependent selection can result in long-term fluctuations in allele frequencies, so that each species remains polymorphic and is continually adapting to the changing genetic constitution of the other. (This would be a clear instance of the Red Queen principle.) Populations of Australian wild flax differ in the frequencies of phenotypes resistant to various strains of flax rust, and associated rust populations likewise vary in virulence. The variation among populations may be an effect of Red Queen dynamics.

B. Selection for Sex

The supposition that parasite populations continually evolve has suggested to many authors that a major advantage of recombination and sexual reproduction may reside in the ability to generate novel, parasite-resistant genotypes. In a species of freshwater snail, for example, sexually reproducing females are more frequent than parthenogenetic females in habitats where the risk of infection by trematodes is greatest. Furthermore, among the several theories of sexual selection by female choice is the proposal that it is advantageous for females to choose males whose elaborate ornaments and behavior indicate that they are not debilitated by parasites, and thus are likely to father parasite-resistant offspring. Evidence from a few studies has been interpreted as support for this hypothesis, illustrating the broad ramifications that coevolution may have.

C. Evolution of Virulence and Avirulence

Often, though by no means always, parasites have less virulent effects on their normal host species than on novel, recently invaded species of hosts. This pattern might result from (1) evolution of resistance in a host species, (2) extinction of highly virulent species of parasites ("species selection"), because they drive their hosts to extinction, and/or (3) evolution within parasite species toward lower virulence. The evolutionarily naive often do not distinguish between the latter two causes, and imagine that parasites evolve avirulence because this will safeguard the host population and thus the perpetuation of the parasite species. However, natural selection is not prescient, and selection within populations can easily lead to fixation of "selfish" genotypes that enhance the risk of extinction. Evolution of the level of virulence depends on many factors.

For a parasite such as a virus, bacterium, or protozoan that reproduces on or in its host, each individual host carries a temporary group of parasites. Consider first a case in which each host carries only a single, clonally propagating parasite genotype. The fitness of this genotype is measured by the number of uninfected hosts it infects, relative to other genotypes. This number is often proportional to the number of offspring produced within the host and capable of transmission. But greater numbers of parasites take a greater toll on a host, resulting in greater virulence. The fitness of a parasite genotype is thus set by a balance between the rate of reproduction within the host and the reduction in transmission due to death or debilitation of the individual host. If a parasite population evolves less than maximal virulence, it is only because this enhances transmission of progeny to new individual hosts, not because it preserves the host population for future use.

The optimal level of virulence depends partly on the mode of transmission, especially whether it is mostly vertical (from parent host to offspring) or horizontal (as in sexually transmitted parasites). The transmission rate of a vertically transmitted parasite is directly proportional to the number of offspring of its carrier, so such parasites should be relatively benign. However, a horizontally transmitted parasite, given opportunity for transmission, does not profit from its host's further survival or reprodtion. Many such parasites kill their hosts (e.g., baculoviruses of insects) or castrate them (e.g., some plant-pathogenic fungi).

The evolution of virulence is still more complex if individual hosts are typically coinfected by multiple genotypes of the parasite. *Within* each host, the genotype with highest growth rate *r* (and highest virulence) has an advantage, but if individual hosts bearing this genotype die before transmitting it, groups of parasites dominated by less virulent genotypes will transmit the most offspring. Thus group selection favors lower virulence, and individual selection favors higher virulence. At equilibrium, the level of virulence is likely to be the higher than in the case of singly infecting parasites. Comparison of the virulence of species of nematodes that infect fig wasps either singly or multiply has supported this theory, and phage in laboratory cultures evolved lower virulence if they were vertically than if they were horizontally transmitted (Bull, 1994).

VII. SPECIAL FEATURES IN HERBIVORE–PLANT COEVOLUTION

A. Resistance, Tolerance, and Overcompensation

Herbivore impact on plant fitness can be intense, imposing selection for traits that reduce the effects of herbivory. One class of traits simply reduces the incidence of herbivory by escaping herbivores in space and/or time, that is, by decreasing the plant's "apparency." The idea is that abundant, long-lived plants in communities of low diversity will be more easily tracked down and fed upon by herbivores than emphemeral plants that establish short-lived populations in more diverse communities. Size of the plant may also be a factor, as small plants with simple architecture generally have lower insect diversity, perhaps by offering little protection to insects. Consistent with this hypothesis is the observation that among Capparales in Morocco, less damage by pierid butterfly larvae is incurred by species that fluctuate more in density. This suggests that greater constancy in population size and density, that is, greater predictability and apparency, increases the likelihood of herbivory. Fluctuations in density could, however, be a result of adaptation of germination to specific environmental cues or other factors rather than pressures from herbivory.

Most plants must deal with the presence of herbivores, and possess a variety of mechanisms that either reduce herbivory (*resistance*) or ameliorate its effects (*tolerance*). The ability to tolerate damage, that is, to suffer a reduced impact of damage on fitness, depends both on extrinsic factors, such as resource availability, the timing of herbivory, and the parts affected, and on characteristics of the plant. Thus it is thought that plants with intrinsically higher growth rates, storage capacity, or higher number of active meristems will have greater opportunity for regrowth in face of biomass loss to herbivory. It is as yet unclear whether these traits are adaptive defenses or whether they evolved because of other selective forces, such as competition and resource assimilation.

Some experiments have suggested that under ideal conditions, damage at certain phenological stages and the presence of abundant resources could render plants more fit than had they not incurred damage, that is, they would overcompensate. For example, artificially damaged scarlet gilia (*Ipomopsis aggregata*) produced approximately twice as many flowers and fruits as undamaged plants. The root biomass of clipped plants was greater than that of controls, suggesting that by increasing root size plants were able to take up the extra nutrients necessary for the overcompensation response. This implies that herbivores would benefit plants by feeding on them. There is controversy over the design and results of these experiments, and overcompensation has not been observed in enough species to be considered a general phenomenon.

Greater attention has been given to plant resistance, namely, to mechanisms that actively repel herbivores or that make the plant unpalatable or undesirable. Resistance mechanisms differ with respect to when they are expressed: many defenses are constitutive, that is, expressed regardless of herbivore presence, whereas others are absent until induced in the presence of herbivores. Induced responses curtail the costs associated with defenses, as these are incurred only when strictly necessary.

Some plants have morphological defenses, such as hard seed coats, thorns, or stinging hairs. However, a central role in defense has been attributed to secondary metabolites. These chemicals play no part in primary metabolism and groups of compounds are usually restricted to a few plant families. They have been found to affect herbivore behavior by decreasing a herbivore's feeding rate, deterring oviposition, slowing development, and reducing fecundity. Yet many of these compounds have also been found to affect bacteria, fungi, viruses, other plants, and other herbivores. Secondary compounds might just as well be an adaptive defense to these organisms, or even serve other ecological or physiological functions (e.g., UV protection, drought tolerance). Convincing evidence that insects are a driving force in the evolution of plant chemical defenses comes from studies that compare the pattern of natural selection on putative resistance factors in the presence and absence of herbivores. For example, in the presence of herbivores a population of *Arabidopsis thaliana* was shown to be under selection for increased levels of mustard oils (or glucosinolates) and trichome densities.

In contrast to resistance, plant tolerance does not affect the performance or fitness of the herbivore. Herbivores will not be under selection pressure to overcome this plant defense and so coevolution should not be expected, unless resistance and tolerance are correlated. Theoretical models do indeed predict a trade-off between these two forms of defense, but the limited evidence so far is inconclusive.

B. Specialist and Generalist Herbivores

Most herbivorous vertebrates and many insects are generalists, feeding on plants in many families. A more specialized diet, of plants in only a single family, genus, or even species, characterizes the majority of herbivorous insects and a few vertebrates (such as the koala, a specialist on *Eucalyptus* foliage). The advantages of a broad diet seem clear: an individual can feed despite spatial or temporal variation in the availability of any one food type. Theoretically, the possible advantages of specialization are clear, but evidence is often more ambiguous.

Commonly, closely related species of insects specialize on different plants. Assuming their common ancestor fed on one or more of these plants, the problem is to explain why some of its descendants, instead of expanding their diet, abandoned their ancestral host to which they were presumably well adapted in favor of new hosts. Several ecological advantages of such a switch have been suggested. The new host might provide escape from predators or parasites. Certainly specialization is maintained, if not originally caused, by the protection that many insects, such as milkweed-feeding monarch butterflies, gain by sequestering defensive plant compounds. Competition for resources, as we have noted earlier, may also select for specialization. In many species, mating occurs only on the host plant, and use of a single host may be the most reliable way to find mates. Host specialization will then be a consequence of positive frequency-dependent selection. There is indirect evidence for each of these hypotheses.

The most general explanation for specialization is likely to be trade-offs in the ability to find, handle, digest, or detoxify different kinds of plants. The secondary compounds that characterize higher taxa of plants are thought to play an especially important role, since many are toxic and/or deter feeding by nonadapted insects. Conversely, host-specific insects typically use some such compounds as feeding or oviposition stimuli. The hypothesis that related specialist insects usually feed on related plants because of the plants' chemical similarity has been supported by a study of *Blepharida* flea beetles: the beetles' phylogeny is more closely mirrored by chemical similarity than by phylogenetic affinity among the species of their *Bursera* host plants.

Tests for trade-offs have had mixed results. There is little evidence that closely related generalist and specialist species differ in their efficiency of digestion of plant species that they both naturally feed on (Futuyma and Wasserman, 1981). There is some evidence that specialist species can find and recognize host plants faster than generalist species, owing to properties of the nervous system, and abundant evidence that closely related species are physiologically better adapted to their own than to each other's host plant species. However, since such differences may have evolved after these species became specialists, most research on trade-offs has compared genotypes within species. Only a few such tests have found clear evidence of trade-offs; different clones of pea aphids, for example, have high fitness on either pea or alfalfa. In no case is the biochemical or physiological basis for such trade-offs fully understood.

VIII. MUTUALISM

In a mutualistic interaction, individuals of each of two species obtain a fitness benefit from the interaction between them. Long-term physical associations are referred to as symbioses. These include endosymbioses, in which microorganisms such as bacteria reside within their hosts. In some cases, endosymbionts have become, in effect, part of the host. Mitochondria and chloroplasts, for example, evolved from endosymbiotic bacteria early in eukaryote evolution.

Some mutualisms have doubtless evolved from casual, fortuitously beneficial interactions; such may be the origin, for instance, of the nectar glands of many plants that attract generalist ants, which attack herbivores and other intruders. Some such mutualisms have become obligate, as in the case of some *Pseudomyrmex* ants that inhabit and defend only certain species of *Acacia* trees that provide food bodies and special domiciles. In many cases, especially symbioses, mutualisms have arisen from parasitic interactions; for example, a complete spectrum from parasitism to mutualism is found in endophytic fungi (Clavicipitaceae). The conditions that favor evolution of parasitism into mutualism are much the same as those favoring the evolution of avirulence (see earlier). Specifically, the fitness of vertically transmitted symbionts depends on the fitness of their hosts, so a benefit provided to the host will enhance the symbiont's fitness. Even without vertical transmission, long-term (e.g., lifelong) association between individual partners satisfies the conditions for "reciprocal altruism." The survival and reproduction of a colony of *Acacia*-dwelling ants depends on the health of its host tree, so it pays to defend it, just as the fitness of the tree depends on nurturing its defending ant colony.

Many mutualisms are rather nonspecific, in that at least one of the species of partners holds a mutualistic relationship with several or many species. Most mycorrhizal fungi, for example, are not highly host-specific, and most species of pollinating or seed-dispersing animals visit many species of plants. There are, however, some exceptions, such as the fig wasps, each species of which develops as a seed predator of one of the several hundred species of fig and is the sole pollinator of that species. The vertically transmitted bacterial endosymbionts (*Buchnera*) of aphids are so intimately tied to their hosts that the phylogeny of the bacteria mirrors faithfully that of the aphids. Unique associations of lineages of aphids and bacteria have persisted for more than 100 million years.

Mutualism can generally be viewed as reciprocal exploitation. Mutualists do not provide benefits to their partners out of altruism; the benefit is either wrested from them (e.g., pollen consumed by pollinating bees) or it is a "payment" that ensures a reciprocal benefit. Thus potential conflict pervades most mutualisms, for selection often favors "cheater" genotypes that do not repay the benefit they receive. For instance, bees often rob nectar by chewing through the bases of flowers, so that they do not pollinate. Many orchids deceive pollinating insects, providing neither nectar nor consumable pollen, but instead produce scents that mimic the insect's sex pheromone and induce "pseudocopulation" by male insects that effect pollination. Some orchids produce no reward whatever. Selection for cheating can destabilize a mutualistic relationship. For example, yucca moths (*Tegeticula*) oviposit in the ovaries of *Yucca*, where the larvae feed on some of the developing seeds. Most species of *Tegeticula* stereotypically (and "carefully") place pollen on the stigmas of flowers in which they lay, and so both the host plant and the moth's offspring benefit. However, several lineages of "cheater" *Tegeticula* have evolved from pollinating ancestors; they do not pollinate, but lay eggs in flowers that have been pollinated by those species that do.

IX. MAJOR CONSEQUENCES OF COEVOLUTION

If it is difficult to document and characterize coevolution between pairs or among small numbers of species, it is harder still to demonstrate the effects that coevolution has had on the properties of communities, ecosystems, or the history of biological diversity. For the most part, questions on so grand a scale have, so far, only tentative or even speculative answers.

A. Phenotypic Diversity

Although no quantitative data are available, coevolution, or at least adaptation to interspecific interactions, has unquestionably enhanced the phenotopic diversity ("disparity") among species. The diverse trophic structures displayed by related species of fishes, birds, and other animals partly reflect adaptation to the characteristics of the different arrays of prey species on which they feed, and partly arise from selection due to interspecific competition, resulting in divergence and specialization in resource use. The diverse morphologies in such adaptive radiations as the African lake cichlids illustrate that ecological and morphological diversification of a clade may be promoted by competition among its member species, but may be constrained by competition from members of other clades, until such competition is relieved.

Predation and parasitism have doubtless enhanced phenotypic diversity among prey. There is much circumstantial support for Ehrlich and Raven's hypothesis that chemical diversity among plant taxa has been driven by herbivores (and, we would add, pathogens). Intraspecific genetic diversity at major histocompatibility loci has almost certainly been maintained by selection imposed by pathogens, and the extraordinarily complex immune system of vertebrates is testimony to the potent selection that pathogens generate. The Mesozoic evolution of shell-crushing predators called forth an immense proliferation of defensive features in the shells of molluscs.

The cooperation and conflict inherent in mutualism have likewise had immense effects on phenotypic diversity. The mycetocytes of aphids, the light organs in some fishes and cephalopods, and the root nodules of legumes are among the many special structures that house symbiotic bacteria. The phenotypic diversity of the Orchidaceae, the largest family of plants, lies largely in the extraordinary variety of flower forms that attract and manipulate diverse pollinating animals.

Although such examples provide undeniable evidence that biotic interactions have had great evolutionary consequences, phenotypic disparity has seldom been measured. Moreover, although the role of interactions in these examples is obvious, the importance of reciprocity (the "co" in coevolution) is not. Rigorous tests of the effects of coevolution remain to be done.

B. Species Diversity

Variation in species diversity among clades, communities, ecosystems, and biomes and through geologic time is affected by many factors. The relative importance of coevolution on rates of speciation and extinction, which together determine diversification rate, is unclear. Only in exceptional cases, such as the divergence of figs and their species-specific agaonid pollinators, is it likely that coevolution directly induces speciation, that is, reproductive isolation. It has often been suggested that animal-mediated pollination and seed dispersal may enhance rates of plant speciation by establishing reproductively viable isolated populations. However, species diversity of angiosperms is only slightly greater in animal- than wind-pollinated families, although it is greater in families that include both biotic and abiotic seed dispersal than in families that use one mode exclusively.

In principle, extinction rates can be increased or decreased by coevolution, although either is very difficult to document in practice. If coevolution has effected a net increase in diversification rate, it has probably done so chiefly in three ways: (1) by shifts in resource use, thus reducing competition; (2) by evolutionary escape from predation, followed by radiation; and (3) by evolution of specialized predators, parasites, and mutualists following the evolution of new prey or hosts. That is, diversification of one group of organisms begets diversification of others, as Ehrlich and Raven postulated for plants and herbivorous insects.

These hypotheses are supported by some modest evidence. Farrell *et al.* (1991) found that clades of plants with canals that deliver defensive latex or resin to sites of herbivore attack are consistently more diverse than their sister groups that lack these features. Similar sister-group comparisons among insects showed that the evolution of herbivory has consistently been followed by greater diversification. However, it is not known if the diversity of herbivorous insects has been enhanced in clades that have adapted to chemically novel, diverse plant clades, as Ehrlich and Raven proposed. In western Europe, the species richness of aphids and certain other insects that feed on different plant families is correlated with the number of species of plants in the family. Whether or not such a correlation holds on a global scale is not known.

C. Community Structure

The convergent evolution of ecologically "equivalent" species, such as many Australian marsupials and their placental counterparts in other regions, suggests that independently evolving communities might converge toward similar structure. Certainly, properties such as vegetative morphology and the architecture of vegetation converge; for example, "Mediterranean" vegetation (chaparral, matorral, maquis) is dominated by sclerophyllous, small-leaved shrubs in several parts of the world. Such similarities, however, arise simply from each species' independent adaptation to physical environmental factors. Whether or not community properties such as stability, food-web structure, or species diversity converge as a result of coevolution is a different, much more difficult question. Simple models of interspecific competition and of food webs suggest that the coevolutionarily stable equilibria might be rather few, so that coevolution might yield some predictable structure (MacArthur and Levins, 1964). Whatever the ideal applicability of these models might be, however, the opportunity for such pervasive effects of coevolution may generally be rather slight. Gene flow among conspecific populations that interact with different ensembles of species may prevent finely tuned coevolution. Moreover, paleoecological studies have shown that throughout the Pleistocene, species have had highly individual histories of change in geographic distribution, so that many of today's assemblages of species are very recent. Except for specialized associations, as of host-specific parasites that have moved about with their hosts, there has been little time for coevolutionary adjustments in many of today's species assemblages. Nevertheless, paleontologists have documented rather steady levels of diversity at both global and local levels over vast periods of time (10^7–10^8 years), despite turnover of taxa (Jablonski and Sepkoski, 1996). These observations suggest that whatever convergence or constancy of community structure exists may be attributable more to purely ecological rules of community assembly rather than to coevolution.

An unusually clear example of convergent multispecies assemblages is provided by the anoline lizards mentioned earlier, which form monophyletic groups of morphologically and ecologically equivalent species on each of the Greater Antilles islands. However, similar habitats in different parts of the world generally are not very similar in species richness. For example, lizard diversity in Australia exceeds that in corresponding habitats in southern Africa. Nevertheless, in some instances variation in species richness among habitats shows similar patterns; lizards are more diverse in deserts than in wetlands in both Australia and Africa. Thus, habitats seem to have consistent effects on species coexistence, and perhaps on the evolution of resource partitioning (Ricklefs and Schluter, 1993).

Coevolution may also affect community properties such as stability, that is, the property of returning to equilibrium after a disturbance. The study of artificially constructed food webs has identified some patterns of interaction among component species that are conducive to stability. One such aspect is connectance, that is, the ratio of actual to potential interspecific links. High levels of connectance are associated with reduced stability, as a disturbance to one species will affect many other species in the community. Thus processes that reduce connectance, or generate compartments within a community, such as a subset of species that interact predominantly with one another, contribute to stability. To the extent that coevolution between parasites and hosts and between plants and herbivores leads to increased levels of specialization (and thus compartmentalization), it can thus indirectly enhance community stability.

See Also the Following Articles

COMPETITION (INTERSPECIFIC) • DARWIN, CHARLES • EVOLUTION, THEORY OF • PARASITISM • PLANT-INSECT INTERACTIONS • RECOMBINATION • SPECIES COEXISTENCE

Bibliography

Abrams, P. A. (1990). The evolution of anti-predator traits in prey in response to evolutionary change in predators. *Oikos* **59**, 147–156.

Bull, J. J. (1994). Perspective—Virulence. *Evolution* **48**, 1423–1437.

Farrell, B. D., D. Dussourd, and C. Mitter. (1991). Escalation of plant defenses: Do latex and resin canals spur plant diversification? *Am. Nat.* **138**, 881–900.

Futuyma, D. J., and M. Slatkin. (1983). *Coevolution*. Sinauer Associates, Sunderland, Massachusetts.

Futuyma, D. J., and S. S. Wasserman. (1981). Food plant specialization and feeding efficiency in the tent caterpillars *Malacosoma disstria* and *Malacosoma americanum*. *Entomol. Exp. Appl.* **30**, 106–110.

Jablonski, D., and Sepkoski, J. J. (1996). Paleobiology, community ecology, and scales of ecological pattern. *Ecology* **77**, 1367–1378.

MacArthur, R. H., and R. Levins. (1964). Competition, habitat selection and character displacement in a patchy environment. *Proc. Nat. Acad. Sci.* **51**, 1207–1210.

Ricklefs, R. E., and D. Schluter (eds.). (1993). *Species Diversity in Ecological Communities*. University of Chicago Press, Chicago.

Taper, M. L., and T. J. Case. (1992). Coevolution among competitors. *Oxford Surveys in Evol. Biol.* **8**, 63–109.

Thompson, J. N. (1994). *The Coevolutionary Process*. University of Chicago Press, Chicago.

Vermeij, G. J. (1987.) *Evolution and Escalation: An Ecological History of Life*. Princeton University Press, Princeton, New Jersey.

COMMONS, CONCEPT AND THEORY OF

Colin W. Clark
University of British Columbia

GLOSSARY

bionomic equilibrium In a biological common property resource, an equilibrium between resource production rate and exploitation rate, characterized by the dissipation of economic rents.

commons (or common property resource) Any resource asset that is not privately owned and controlled.

discount rate Rate (usually annual) at which future revenues are discounted to calculate a present value. Personal discount rates often exceed market rates by a wide margin.

economic rent The flow of net economic benefits (revenue minus costs) derived from the exploitation of a resource asset.

externality Cost or benefit imposed on others as the result of some economic activity.

Pareto efficiency An economic equilibrium is Pareto efficient if any change will decrease the economic well-being of at least one participant.

tragedy of the commons Process whereby a commons is overexploited because individual users cannot expect to realize the potential benefits of resource conservation.

IN MEDIEVAL ENGLAND THE TERM "COMMONS" referred to an unfenced area of land that, although privately owned (usually by the local manor lord), was by tradition available to specific persons (the commoners) for specific activities, such as pasturage or removal of building materials. The subsequent enclosures of the commons, for the sole benefit of their owners, reduced many former users to penury. In recent times, and especially following the publication of the famous article "The Tragedy of the Commons" (Hardin, 1968), the term has been used to refer to any resource asset that is not privately owned and controlled. This is the sense in which the term commons, synonymous with "common property resource," will be used here.

Encyclopedia of Biodiversity, Volume 1

I. INTRODUCTION

Important contemporary examples of common property resources include the global atmosphere, the oceans, large lakes, rivers, forests, and fish and wildlife populations, including birds. Though not inevitable, the overexploitation of common property resources is always a potential threat, and often a frightening reality. Many current environmental problems can be traced to the working of a so-called tragedy of the commons.

The classification of resources as either common or private property is an oversimplification. Many gradations exist between a total lack of access restriction (i.e., open access) and complete individual control. For example, coastal marine resources within 200-mile fishing zones are now recognized as being under the sole jurisdiction of the coastal state, with foreign fishermen either excluded or subject to payment of fees. These resources are thus limited access, but (unless privatized) still common property by our definition. Also, private landowners seldom have full rights to do as they wish with their property—the state may retain mineral rights, for example, and zoning regulations may restrict land use and development.

The logic of the tragedy of the commons is straightforward, as will be explained later. Unfortunately, however, much of what has been written about common property resources suffers from vagueness and imprecision, leading to unnecessary confusion and controversy. The following hypothetical example can be useful as a basis for a general understanding of the economics of a commons.

Imagine a small lake, completely surrounded by the properties of two landowners. The lake contains a population of fish, and both landowners are avid fishermen. What will be the long-term outcome? The answer is not obvious and depends on various factors.

The first possibility is that the lake is so large, and the fish population so productive, that no amount of fishing by the two landowners has any appreciable effect on the population. In this ideal situation, by assumption nothing happens—the tragedy of the commons does not occur.

The opposite case is more instructive. Suppose, for example, that upon first taking up their properties, the two neighbors are able to catch all the fish they want. After a few years, however, they observe that fish are becoming more scarce. Also, the big ones that used to put up such a fight seem to have disappeared. Subsequent developments now depend on how the two landowners react. If they meet on amicable terms and agree to restrict their catches, or to return some of them to the lake, the dynamics of the exploited fish population may stabilize, permitting sustained if limited catches by both owners. It may take several years of trial and error before the maximum sustainable catch is determined—or they may decide to forego the maximum catch in favor of maintaining a large population of large fish.

An alternative solution might be to subdivide the lake into two fishing zones, one for each landowner. But this assumes cordial relations between the landowners, with each cooperating in not fishing in the other's zone. If fish are able to swim freely between zones, such an agreement may fail to prevent overfishing. Privatization of property or resources is not always feasible, and when it is feasible it presumes mutual trust and cooperation, or external enforcement.

Instead of such amicable agreements, however, the owners may begin blaming each other for overfishing. Relations may then deteriorate, with both determined to catch all the fish they can find. One may buy a faster boat, hoping to outdo his neighbor, who then responds in kind, and so on. Such behavior between two neighboring individuals may seem petulant. But what if instead of two property owners, there were 100 cottages on the lake, all owned by avid fishermen? Besides increasing the pressure on the fish population, this would make the problem of agreeing over and enforcing catch restrictions much more difficult. Cheating by individuals would be hard to detect, and even harder to control.

This simple example illustrates many of the characteristic features of common property resources, including potential overexploitation, the necessity of mutual cooperation, and the increased difficulties of control as the number of joint owners increases. We return to these questions later.

II. THEORY

The theory of a common property renewable resource will be discussed here in terms of a specific example, a commercial fishery. The results and predictions apply, however, to most other biological resources, with minor modification.

The theory of bionomic equilibrium in an unregulated, open-access, common property fishery was initiated by Gordon (1954)(Fig. 1). Fishermen exert fishing effort E on a certain fish population; fishing effort is a measure of fishing intensity, for example, the number of vessel-days per year, or (in a rod-and-reel fishery) the number of rod-hours per year. Figure 1 shows the

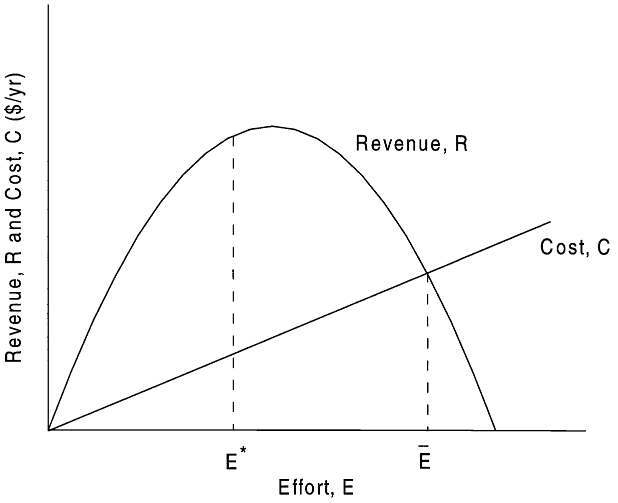


FIGURE 1 Gordon's (1954) model of equilibrium revenue *R* and cost *C* in a fishery; $\overline{E}$ = bionomic equilibrium; E^* = effort level for maximum sustainable economic rent.

total sustained annual revenue *R* and total cost *C* as functions of effort *E*. Cost increases in proportion to *E*, but revenue rises to a peak ("maximum sustainable yield") and then declines at higher levels of effort. An input–output function having this inverted-U shape is unconventional in economic theory; it reflects the underlying biology of the resource, which becomes increasingly unproductive as it is depleted by high levels of fishing effort.

Gordon's prediction is that the fishery will reach an equilibrium, called the bionomic equilibrium, at the effort level $\overline{E}$ where revenues and costs are equal. The argument is as follows. First, for $E < \overline{E}$, revenue exceeds costs and fishermen are making money in excess of their opportunity costs. This positive net revenue attracts additional entrants to the fishery, resulting in an increased effort level. On the other hand, if $E > \overline{E}$, fishermen are losing money and some will leave the fishery. Equilibrium can only occur where $E \equiv \overline{E}$. At this bionomic equilibrium the fishermen are barely breaking even financially. (See Clark, 1990, for mathematical representation of these ideas.)

For levels of *E* below $\overline{E}$, revenues *R* exceed costs *C*; the excess $R - C$ is called economic rent. Gordon's prediction can then be stated: in an open-access unregulated fishery all potential economic rents will be dissipated in long-run bionomic equilibrium. Additional predictions are that if the price of fish increases, or if the cost of fishing decreases, the effort level $\overline{E}$ will increase, resulting in a more depleted fish population. What would normally be recognized as benefits to the fishing industry (higher prices, lower costs) are dissipated because of the tragedy of the commons.

What is not clear in the Gordon model is whether the fish population will survive in the long run, or be driven ultimately to extinction. In practice, very few fish species have been fished to extinction, although some marine mammals have been. On land, America's most abundant bird species, the passenger pigeon, was hunted to extinction in the nineteenth century. Today species such as rhinoceroses, tigers, and certain birds are endangered by market hunters. If we agree that the Gordon theory applies to any such commons, then any species is likely to become endangered if the value of the last specimen caught exceeds the cost of capturing it. In a world of extremely affluent consumers and destitute hunters, this condition applies to an ever-widening class of wild species, both marine and terrestrial.

Gordon also identified the effort level E^* yielding maximum economic rent (see Fig. 1). He pointed out the apparent anomaly that if the fishermen would only work less hard—reduce their effort—they would be better off! This argument may seem compelling, but the reality is that many, if not most, marine fisheries today remain closer to bionomic equilibrium than to rent maximization. It often seems as if the fishing industry is determined to suffer the tragedy of the commons even though impoverishment is the inevitable outcome. What explains this phenomenon?

To address this question we must consider the dynamics of the fish population (recall that Fig. 1 pertains only to equilibrium situations). For example, suppose that a certain fishery is currently at bionomic equilibrium $\overline{E}$. A reduction of effort is needed in order to improve the situation. Unfortunately, such a reduction of effort will not lead to an immediate increase in catches—quite the reverse. At first catches will decrease, more or less in proportion to the decreased effort. The decreased catches will (it is hoped) allow the fish stock to increase over time. Eventually a new equilibrium will be reached, combining reduced effort with increased catches and increased rents. A simulated illustration of this process is shown in Fig. 2.

The principal message from the argument is that resource conservation necessarily involves a sacrifice of potential current benefits in order to achieve greater future benefits. From the economic point of view, conservation is thus a form of investment in the future.

We also see now why management of a common property resource is likely to be difficult: the resource users must agree to make the necessary sacrifice (investment). Moreover, all resource users must then somehow be prevented from overexploiting the protected re-

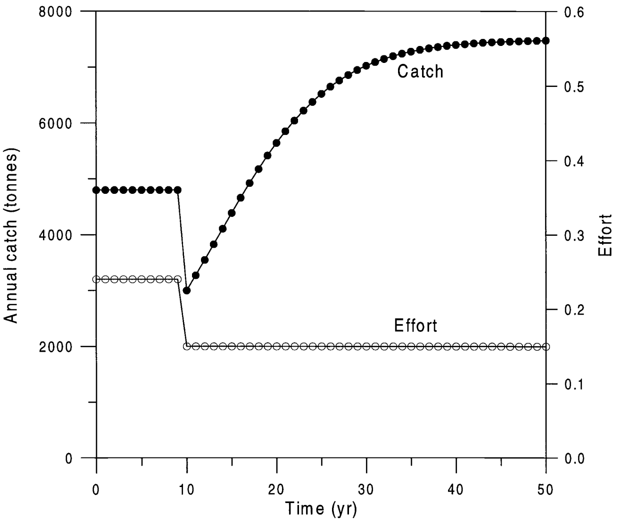


FIGURE 2 Simulated annual catches following a reduction in effort at time $t_0 = 10$/yr. At first catches are reduced, but eventually catches increase to a higher level than before, as the fish population recovers from overharvesting. (Discrete logistic model; $r = 0.3$/yr, $K = 100{,}000$ metric tons, $E_{t<t0} = 0.24$/yr, $E_{t>t0} = 0.15$/yr.)

source, even though such activities would be individually profitable. The more successful the management program, in terms of protecting the resource, the greater will be the incentive for excessive exploitation by individuals. This dilemma is the real tragedy of the commons.

Although formulated here in terms of fish populations, the concepts of bionomic equilibrium and rent dissipation in common property resources are widely applicable. Overgrazing of grasslands, desertification, deforestation, soil erosion, and pollution are other important examples of overexploitation of the commons. Everywhere that resource stocks are used as common property, the incentive for the commons tragedy exists.

The participants in a tragedy of the commons are sometimes said to be motivated by short-term financial considerations—with "everyone following their own short-term interest" to quote from a recent publication. This portrayal of such resource users as greedy, short-sighted souls is often both unfair and inaccurate (although there probably is a tendency for such individuals to gravitate toward unregulated resource industries). In a commons situation, it is in each participant's best interests to get as much as possible out of the resource before it is depleted by others. Any personal sacrifice for the sake of resource conservation is simply dissipated by the other users.

III. REGULATED BIONOMIC EQUILIBRIUM

How can the problems of common property resource exploitation be overcome? Let us first consider an approach that often does not, and perhaps cannot, succeed. Again, imagine a fishery initially at bionomic equilibrium (this is not essential for the discussion). A government agency is set up to manage the fishery. Realizing that the problem is excessive effort in the fishery, the managers introduce a limited-entry program, in which only a fraction of the previous participants are licensed to catch fish. (It goes without saying that, unless compensated in some way, the excluded fishermen will tend to resist this program.) At first the new system may appear to be successful—effort decreases, the fish population begins to recover, catch rates go up; the remaining fishermen are making money.

But now the economic dynamics of the commons begin to operate once again. By increasing their own levels of effort, individual fishermen can increase their current income. Some of the fishermen buy bigger boats, forcing others to do the same in order to compete. The managers must now bring in regulations to restrict the size of fishing boats. The owners then respond by "capital stuffing"—installing larger engines, more powerful winches, better electronic devices, and so on, all of which increase the fishing power of their vessels. These items in turn must then also be regulated. Eventually either the regulations become so cumbersome that economic rents are dissipated simply by the complexity of the management system, or else the regulations fail to prevent overexpansion of fishing capacity. In either case, the fishery is now in regulated bionomic equilibrium.

Such regulated bionomic equilibrium can nowadays be seen in many managed marine fisheries. Where the regulations are effective and strongly enforced, depletion of fish stocks may be prevented. In practice, however, the economics of the commons has often proved more powerful than the regulations. In a number of cases, intensively managed fisheries have collapsed from persistent overfishing, cod and groundfish on the Atlantic Coast of the United States and Canada being notorious examples. In most of these cases, the fishing industry itself persistently and adamantly opposed the imposition of the regulations that were required to preserve the fishery.

While all of this may seem quite irrational (unless you understand the economics of the commons), in actuality the situation is often more complicated—and more perverse—than described here. To be specific, resource industries are often highly subsidized. Individuals can only benefit from the subsidy if they participate in exploiting the resource. The effect of the subsidy is to reduce the opportunity cost *C*, in extreme cases to near zero. The Gordon theory explains what happens: bionomic equilibrium corresponds to virtual resource extinction in such a situation. Exploiters become dependent on the subsidy, not the resource. Bionomic equilibrium is more devastating with subsidies than without them.

In the case of Canada's Atlantic cod fishery, the system of regulated, subsidized bionomic equilibrium proceeded to the ultimate conclusion with the almost complete disappearance of cod stocks in 1991. This collapse destroyed the basis of the economy of the entire province of Newfoundland; since 1991 the Canadian government has been paying over $1 billion per year as welfare to displaced Newfoundland fishermen and plant workers. Whether the Newfoundland cod will ever return is uncertain at the present time (1999).

IV. MANAGING INCENTIVES VERSUS MANAGING RESOURCES

Two general approaches to the management of common property resources can be identified: those directed toward management of the resource stock and those that alter the incentives of resource users. Methods aimed at managing the resource include total annual catch quotas, controlled access to the resource, restrictions on the technology of harvesting the resource, and so on. Methods that rely on altering the economic incentives of common property resource exploiters include user fees or royalties and individually allocated quotas (see the following). These methods may also be used in combination. As indicated by the foregoing discussion, management strategies that do not alter users' basic incentives tend to result in regulated bionomic equilibrium, with excessive inputs and persistent rent dissipation. Methods that alter incentives have the potential of overcoming these difficulties, although the persistent incentive to cheat must be recognized, and countered by monitoring and strict enforcement of the rules.

A third approach, privatization of the resource—usually involving geographical subdivision into individually manageable units—may be considered the ideal solution, since it allows users to become independent of each other and of excessive government interference. This is historically the usual method for dealing with nonfugitive resources, such as land and minerals. Such subdivision and privatization is not feasible, however, for many of today's remaining common property resources. In such cases, some form of centralized management seems necessary. We first consider the case in which the resource is located within the jurisdiction of a single state; the case of international common property resources is discussed in Section VI.

V. ALLOCATED QUOTAS

One method of altering the incentives of resource users is a system of allocated quotas, often sold under auction to the highest bidders. This method is commonly used in logging in national forests and in offshore oil drilling.

Having obtained such a quota, the users' incentives are to harvest the quota at minimum cost. Persistent incentives toward cheating are thwarted by rigorous enforcement of the quotas. Resource rents are preserved, and divided between resource users and the government. The resource remains common (state) property, but specific user rights are privatized through the system of quotas. Multiple uses of the resource, as in national forests used for both timber production and recreational activities, can be achieved, although different user groups may attempt to influence management decisions in their favor.

The establishment of 200-mile national fishery zones in the 1980s rendered coastal marine resources subject to quota-based management systems, and individual catch quotas are indeed now in use in many such fisheries. Marine populations, however, provide unusual management difficulties, not the least of which is the fact that these resources are largely hidden from view until they are brought to the surface by net or hook. Also, fish populations undergo major natural fluctuations, which are poorly understood and hard to predict. In addition, many fish populations migrate, with migration routes that change in response to oceanic conditions. These features introduce large, irreducible levels of uncertainty into fisheries management; we discuss this topic further in Section IX.

VI. INTERNATIONAL COMMON PROPERTY RESOURCES

As an important example of an international common property resource, let us briefly consider the atmosphere. Although partitioning of the atmosphere is clearly not possible, policies at the national level have had significant effects on atmospheric pollutant emissions. Norway, Sweden, Denmark, France, and Japan have employed taxes on the emission of sulfur dioxide since the 1970s. Some countries now use tradable quotas to control atmospheric emissions. Such policies affect incentives, and are thought to be far more efficient than direct quantitative regulations.

At the international level, over 100 nations have now ratified the 1987 Montreal Protocol requiring the phasing out of ozone-depleting chlorofluorocarbons and related substances. This example shows that, when evidence of severe harm is strong, international cooperation can be achieved. However, the example also illustrates the difficulties, especially at the international level, associated with common property resource management, as noncompliance, failure to ratify, and the proliferation of illegal trade in ozone-depleting substances have persisted. Although no international agency exists to enforce the Protocol, trade sanctions (or the threat thereof) have been used to encourage adherence to the rules.

Similar agreements, and methods of encouraging nations to follow them, apply to other international resource problems, such as greenhouse gas emissions, marine resource exploitation, ocean pollution, and the conservation of migratory birds. Although international management may lag behind the need, it is encouraging that the general principles of managing the commons do appear to be having an effect on the way in which the rules are being written.

VII. DISCOUNTING THE FUTURE

The view is sometimes expressed that privatizing a commons will automatically result in its conservation. Some economists have gone so far as to assert that the private owner of a renewable resource asset would never overexploit it, except perhaps as a result of miscalculation. This assertion is totally false and is based on a failure to understand the economic dynamics of natural resource stocks.

To summarize the argument in a nutshell (see Clark, 1990, for details), the hypothetical owner of a renewable resource will calculate the total discounted present value of various options, including sustainable harvesting and depleting the resource for an immediate gain. Managing a certain resource sustainably might, for example, produce an annual income I, whereas eliminating the resource stock could yield an immediate gain P. The present value of the perpetual income stream I is equal to I/i, where i is the annual discount rate. If it happens that P is larger than I/i, it will be more profitable to deplete the resource. By investing the proceeds at interest rate i, the owner will obtain a greater income than by harvesting it sustainably.

Many biological resources, such as whales, trees, and soil, have low annual rates of productivity, often less than 5% per annum. Removing more than 5% of the stock each year will deplete the resource, eventually to extinction. If the resource owner discounts future revenues at more than 5% per annum, he or she may deliberately choose to eliminate the resource for short-term gain.

Resource users may have reasons for discounting future income at rates higher than market interest rates. For example, they may need, or desire, an immediate cash flow. In addition, the future of the resource may be uncertain, enhancing the relative value of current harvests. Whatever the reason, a resource owner may decide to practice less conservation than may be consistent with sustainable use of the resource.

The same calculation applies to multiple users of a managed common property resource. It is a notorious fact, for example, that the fishing industry almost always demands larger quotas than recommended by government managers. The managers' calculations may be based tacitly on zero discounting, but the industry almost always prefers to receive its income now.

VIII. COMMUNITY CONTROL

A subsidiary theme in the common property resource literature proposes that common property can be a good thing, not necessarily the stage of a tragedy. The idea is that if all so-called "stakeholders" can be assured a voice in the management of the resource, then agreement to achieve the optimum exploitation should follow. Doubtlessly this could be true in a tightly knit society in which defectors face ostracization or other severe penalties. Recent attempts to reach compromise agreements between users of forests or fisheries, however, have proven frustrating. The hypothetical example of the lake with two adjacent property owners suggests that resource conservation can depend strongly on individual behavior. Although decentralization can sometimes bring benefits, the likelihood that community control of biological resources will improve the prospects for conservation needs to be assessed critically in each instance.

As described earlier, individual community members may discount the future to such an extent that the results differ little from the tragedy of the commons. Also, if there is any likelihood that some users will try to take more than their share of the resource, community control can quickly degenerate into a destructive commons scramble. Always a master of the rapier phrase, Hardin (1994) labeled those who preach the doctrine of community control as "commonists." In some cases, community control might resolve a local tragedy of the commons, but the number of individuals involved has to be fairly small, and all outsiders must be excluded. Furthermore, the community probably needs to be homogeneous in tradition and religion.

IX. RISK AND UNCERTAINTY

Risk and uncertainty are important in the management of many resources, including common property resources. Here, the following terminology is used: Risk occurs when the outcome of some process is not known in advance, but the probabilities of the various possible outcomes are known from past experience. Uncertainty occurs when these probabilities are not known, or not known accurately. Uncertainty can be reduced through experience and research, but in most situations it cannot be eliminated entirely, that is, it cannot be transformed into pure risk. Examples of irreducible uncertainty are common in resource management. To mention one example, we do not know for certain whether current global warming is the result of increased levels of greenhouse gases, nor do we know what impact greater concentrations of these gases will have.

In the case of fishery resources, uncertainty persists regarding current stock levels, relationships between stock and recruitment, and the overall effects of exploitation on future stock levels. Furthermore, marine ecosystems are not well understood, meaning that the implications of harvesting certain species cannot be predicted with certainty. In addition, uncertainty pertaining to future environmental conditions, such as global warming, only adds to the difficulty of foreseeing future trends in marine populations.

Reducing uncertainty (e.g., through scientific research), or reducing the effects of uncertainty (e.g., by adopting conservative harvest strategies), can be considered as forms of investment in maintaining the future viability of a renewable resource. The tragedy of the commons operates here as well—under common property conditions, resource exploiters ignore future uncertainties as they attempt to maximize current revenues from the resource. Management strategies that limit resource exploitation on the basis of scientific uncertainty are widely unpopular with fishermen and other common property resource users. Indeed, the argument is often advanced that uncertainty implies that the need to restrict resource use has not been proved.

X. BIOLOGICAL RESERVES

One response to the ever-growing pressure on biological commons is the establishment of protected biological reserves. While the benefits of reserves are apparent in terms of protecting biodiversity, the potential direct economic benefits of reserves have until recently been less appreciated. For example, large-scale marine reserves could help in preventing severe overexploitation, or collapse of important marine resources. Such marine reserves should be thought of as complementary to, rather than as substitutes for, other types of resource management. Unlike other management strategies, however, reserves can provide a positive, potentially "fail-safe" backup to traditional resource management.

XI. RELATED CONCEPTS

Closely related to the concept of a commons is the economist's concept of externalities (sometimes called spillover effects). An externality is a cost or benefit imposed on others (without compensation) as the result of some economic activity. Externalities can be positive (e.g., a homeowner paints her house to protect it from the weather, and neighbors enjoy the color scheme) or negative (smoke from the homeowner's fireplace chokes her neighbor). Users of common property resources impose negative externalities by reducing the stock of the resource available to other users. Environmental pollution is another common example of a negative externality. In this case, the quality of a resource, rather than the quantity, is reduced by pollution.

Another related economic concept is that of social cost. When a negative externality occurs, the agent that causes the externality incurs a personal cost (his or her private cost) that is less than the total social cost of the activity.

Yet another notion is that of free riders. When the users of a common resource agree to limit their individual impacts, others—the free riders—may continue to exploit the resource. Examples include poachers who illegally slaughter wildlife in protected areas, or nations that fail to ratify or honor international agreements. The dilemma here is that the greater the success in managing the commons, the greater is its attraction to free riders. The problem of free riders explains the dual difficulties inherent in community control of the commons: exclusion of outsiders, and control of cheating by insiders. The failed experiments of national communism serve as strong warnings against undue optimism in these situations.

A basic theoretical prediction of economics, that competitive equilibria are Pareto efficient, is valid only in the absence of externalities. How to remedy the effects of negative externalities is the subject of welfare economics. The immense literature in this field cannot be succinctly summarized, but it can be stated that no fully satisfactory solution to the tragedy of the commons (negative externalities, if you prefer) has been discovered, and it may be that none exists. If not, the commons dilemma promises to become ever more serious as the world's population continues to grow and exert increasing pressures on the biological systems that are essential to our very survival.

See Also the Following Articles

COMMONS, INSTITUTIONAL DIVERSITY OF • ENVIRONMENTAL ETHICS • FISH STOCKS • NATURAL RESERVES AND PRESERVES • POLLUTION, OVERVIEW • PROPERTY RIGHTS AND BIODIVERSITY • RESOURCE EXPLOITATION, FISHERIES • RESOURCE PARTITIONING

Bibliography

Andelson, R. V. (1991). *Commons without Tragedy: Protecting the Environment from Overpopulation—A New Approach.* Barnes and Noble, Savage, Maryland.

Berkes, F. (ed.). (1989). *Common Property Resources: Ecology and Community-Based Sustainable Development.* Belhaven Press, London.

Clark, C. W. (1990). *Mathematical Bioeconomics: The Optimal Management of Renewable Resources.* Wiley–Interscience, New York.

Gordon, H. S. (1954). The economic theory of a common property resource: The fishery. *J. Political Econ.* **62**, 124–142.

Hardin, G. (1968). The tragedy of the commons. *Science* **162**, 1243–1247.

Hardin, G. (1994). The tragedy of the unmanaged commons. *Trends in Ecol. and Evol.* **9**, 199.

Lauck, T., C. W. Clark, M. Mangel, and G. R. Munro. (1998). Implementing the precautionary principle in fisheries management through marine reserves. *Ecol. Appl.* **8**(Supplement), 872–878.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge University Press, Cambridge, United Kingdom.

Shackell, N. L., and J. H. M. Willison. (1995). *Marine Protected Areas and Sustainable Fisheries.* Centre for Wildlife and Conservation Biology, Acadia University, Wolfville, Nova Scotia, Canada.

COMMONS, INSTITUTIONAL DIVERSITY OF

Elinor Ostrom
Center for the Study of Institutions, Population, and Environmental Change and Workshop in Political Theory and Policy Analysis, Indiana University

---

GLOSSARY

common-pool resources Resources that include all ecosystems that are large enough such that excluding potential beneficiaries from their use is a nontrivial task and each individual's consumptive use (e.g., harvesting of a boatload of fish or a truckload of forest products) reduces what is available to others.

governance systems Sets of rule configurations used to govern human–ecosystem relationships at operational, collective-choice, and constitutional levels of analysis.

polycentric governance arrangements Complex, multitiered governance systems in which there is no single center of authority.

rules Commonly understood, normative statements that specify who must, must not, or may take some action or affect some outcome at a particular node in a decision tree.

GOVERNANCE SYSTEMS USED TO REGULATE BIOLOGICAL COMPLEXITY need to be organized at multiple scales and linked effectively together. The design of rules must be viewed as policy experiments with a probability of failure. Highly decentralized governance systems possess both capabilities and limits, as do highly centralized systems. Polycentric systems tend to draw strength from both the lower levels of relatively autonomous units as well as the overarching large-scale units. Institutional diversity is viewed as essential to the preservation of biological diversity.

I. INTRODUCTION

Ecosystems characterized by high levels of biodiversity are complex systems with interactions occurring at multiple spatial and temporal scales. The task of enhancing and sustaining biodiversity through the design of institutions needs to take into account the complexity of the natural systems in which biodiversity occurs. Without effective institutions to limit who can use and regulate diverse harvesting and management practices, com-

Encyclopedia of Biodiversity, Volume 1

mon-pool resources can be overharvested and even destroyed. When common-pool resources containing endemic or threatened species are destroyed, the amount of biodiversity in the world is reduced.

When complexity characterizes the nature of the systems to be governed, it is essential to think seriously about the complexity of the proposed governance systems. Without sufficient concern for sustaining complex governance systems, the very processes of regulating behavior to preserve biodiversity will produce the tragic and unintended consequence of destroying the complexity we are trying to protect. In his book *Design for a Brain: The Origin of Adaptive Behavior* (1960), W. Ross Ashby, an eminent biologist of an earlier era, developed the "Law of Requisite Variety." This law can be roughly stated as: Any regulative system needs as much variety in the actions that it can take as exists in the system it is regulating. Translated into the subject of this Encyclopedia, the law of requisite variety could be stated as: Any governance system that is designed to regulate complex biological systems must have as much variety in the actions that it can take as there exists in the systems being regulated. This is a tall order, but it is one to which we need to pay serious attention.

Unfortunately, much of contemporary policy analysis does not base recommendations on the law of requisite variety. Policy proposals related to the preservation of biodiversity tend to focus on two entirely different solutions: (1) the creation of effective market institutions and (2) the creation of national agencies to manage all common-pool resources within a large domain. Both markets and the state are important institutional mechanisms that should be part of the solutions of protecting biodiversity, but not the exclusive mechanisms. A third option, that of combining large numbers of relatively autonomous local governance systems for specific common-pool resources with larger-scale governance regimes—in other words, a polycentric system—has not been given sufficient attention.

This article focuses on monocentric and polycentric forms of regulation. Section II discusses the underlying presumption of substantial policy proposals, which is that strong, centralized regimes are essential to managing common-pool resources. Substantial empirical work challenges the validity of this presumption. Section III explores the complexity of using rules as tools to change the structure of commons dilemmas, presents the seven clusters of rules that affect the components of any action situation, and then describes the specific rules that are used in field settings by resource users and government agencies. An examination of the types of rules used in the field yields several important findings. First, the *number* of rules actually used in field settings is far greater than generally recognized. Second, the *type* of rules is also different. Boundary rules tend to include as co-appropriators of a resource those who are more likely to be trustworthy because they live permanently nearby and have a long-term stake in keeping a resource sustainable. Authority rules define rights and duties that are easy to understand, directly related to sustaining the biophysical structure of the resource, and easy to monitor and enforce. Some rules recommended in the policy literature are not found among the rules used by self-organized systems. This leads to the difficult search for optimal rules, which is discussed in Section IV.

Given the complexity of the process of designing rules to regulate the use of common-pool resources, I argue in Section V that all policy proposals must be considered as experiments. No one can possibly know whether a proposed change in rules is among the more optimal rule changes, or even whether a rule change will lead to an improvement. All policy experiments have a positive probability of failing. In Section VI, I draw on recent research by John Holland and colleagues at the Sante Fe Institute to discuss the attributes and mechanisms of a different form of general organization—a complex adaptive system—that is not the result of central direction. Complex adaptive systems cannot be understood if one tries to fit these systems into the image of an organization with a central director.

Section VII then shows how the parallel efforts by a large number of local resource users to search out and find local rule configurations may find better rule combinations over the long term, whereas top-down design processes are more limited in their capacities to search and find appropriate rules. All forms of decision making have limits. In Section VIII, I discuss the limits of a series of completely independent resource governance systems and the importance of building polycentric governance systems with considerable overlap to combine the strengths of parallel search and design processes with the strengths of larger systems in conflict resolution, acquisition of scientific knowledge, monitoring the performance of local systems, and the regulation of common-pool resources that are more global in their scope. The resulting polycentric governance systems are also not directed by a single center. They, too, are complex adaptive systems requiring policy analysts to change their fundamental views of organization in order to cope more effectively with tragedies of the commons and many of the other problems facing modern societies.

II. CENTRALIZED REGIMES AND BIODIVERSITY

A substantial number of contemporary policy proposals stress the desirability of central regimes for the regulation of natural resource systems. The "scientific management of natural resources," for example, which is regularly taught in many universities to future regulators of natural resources, presents fisheries and forests as relatively homogeneous systems that are closely interrelated across vast domains. This approach, as it has been applied to fisheries management, is described by Acheson *et al.* (1998: 391–392):

For those trained in scientific management, it is also an anathema to manage a species over only part of its range.... [I]t is not rational to protect a species in one zone only to have it migrate into another area where it can be taken by other people due to a difference in regulations. Lobsters, for example, extend from Newfoundland to the Carolinas.... From the point of view of the National Marine Fisheries Service, it makes sense to have a set of uniform regulations for the entire US coast rather than one for each state.

Contemporary policy analysts also share a belief in the feasibility of designing close-to-optimal rules to govern and manage species over a large domain by utilizing top-down direction. Since natural resources are viewed as relatively homogeneous, and interlinked and simple models exist of how they work, officials acting in the public interest are considered capable of devising uniform and effective rules for an entire region. Prescriptions calling for central governments to impose uniform regulations over most natural resources are thus consistent with much of contemporary policy analysis.

These prescriptions are not, however, supported by extensive research. An extraordinary number of field studies have found that local groups of resource users, sometimes by themselves and sometimes with the assistance of external institutional arrangements, have created a wide diversity of institutional arrangements for coping with common-pool resources when they have not been prevented from doing so by central authorities. These empirical studies document successful self-organized resource governance systems in diverse sectors in all parts of the world. Examples also exist of commons dilemmas that have continued unabated. One conclusion that can firmly be made in light of extensive empirical evidence is that overuse and destruction of common-pool resources is not a determinant and inescapable outcome when multiple users face a commons dilemma. Scholars have begun to identify the conditions of a resource, and of the users of a resource, that are most conducive to local users self-organizing to find solutions to commons dilemmas. Furthermore, the broad design principles that characterize robust self-organized resource governance systems that have resolved commons dilemmas for long periods of time have been identified (E. Ostrom, 1990) and found basically sound by other scholars.

Another important set of findings is that national governmental agencies have been notably unsuccessful in their efforts to design an effective and uniform set of rules to regulate important common-pool resources across a broad domain. Many developing countries nationalized all land and water resources during the 1950s and 1960s. The institutional arrangements that local users had devised to limit entry and harvesting lost their legal standing, but the national governments lacked funds and personnel to monitor the use of these resources effectively. Thus, common-pool resources were converted to *de jure* government-property regimes, but reverted to *de facto* open-access regimes. The perverse incentives of such open-access regimes are often accentuated, since local users have specifically been told that they will not receive any benefits if they adopt a long-term perspective and regulate their use of the resource. When resources that were previously controlled by local participants have been nationalized, state control has usually proved to be less effective and efficient, if not disastrous, than control by those directly affected. The harmful effects of nationalizing forests that had earlier been governed by local user groups have been well documented in Asia and in Africa. Similar results have occurred with regard to inshore fisheries taken over by state or national agencies from local control by the inshore fishermen themselves.

These findings challenge the theoretical foundations of much of contemporary policy analysis. A foundational belief underlying this analysis is that designing rules to change the incentives of participants is a relatively simple analytical task best done by objective analysts not intimately related to any specific resource. Analysts view most resources in a particular sector as relatively similar and sufficiently interrelated, and thus conclude that they need to be governed by the same set of rules. This is bolstered by the view that an organization itself requires *central direction*. Consequently, the multitude of self-organized resource governance systems are viewed as mere collections of individual agents each out to maximize their own short-term returns. The groups who have actually organized themselves are frequently invisible to those who cannot imagine organization without rules and regulations issued by a central authority.

In order to understand why designing rules to regulate the use of common-pool resources so as to sustain their long-term productivity and biodiversity is not a simple task, one has to change perspective somewhat. One has to begin to think through the process of using rules as tools to change the structure of incentives facing users of a common-pool resource.

III. USING RULES AS TOOLS TO CHANGE THE STRUCTURE OF INCENTIVES IN COMMON-POOL RESOURCES

With this change in perspective, we can think of users trying to understand both the biophysical structure of a common-pool resource and how to affect each other's incentives in order to increase the probability of sustainable use over the long term. Instead of being endowed with complete knowledge of how a common-pool resource reacts to use patterns over time, they have to explore and discover the biophysical structure of a particular resource that will differ on key parameters from similar resources in the same region. Further, they have to cope with considerable uncertainty related to the weather, complicated growth patterns of biological systems that may at times be chaotic in nature, and external price fluctuations that affect the costs of inputs and the value of outcomes. In addition to the physical changes that they can make in the resource, seven clusters of rules directly affect the components of their own action situations.[2] Specifically, the rules they can change include:

I. Boundary rules affect the characteristics of the *participants*.
II. Position rules differentially affect the capabilities and responsibilities of those in *positions*.
III. Authority rules affect the *actions* that participants in positions may, must, or must not do.
IV. Scope rules affect the *outcomes* that are allowed, mandated, or forbidden.
V. Aggregation rules affect how individual actions are *transformed* into final outcomes.
VI. Information rules affect the kind of *information* present or absent in a situation.
VII. Payoff rules affect assigned *costs and benefits* to actions and outcomes.

Given the nonlinearity and complexity of common-pool resource situations, it is rarely easy to predict what effect a change in a particular rule will produce. For example, a change in a boundary rule to restrict the entry of users simultaneously reduces the number of individuals who are tempted to break authority rules, but it also reduces the number of individuals who monitor others or contribute funds toward hiring a guard. Thus, the opportunities for rule breaking may increase. Further, the cost of a rule infraction will be spread over a smaller group of users and, thus, the harm to any individual may be greater. Assessing the overall effects of a change in boundary rules is a nontrivial analytical task. Instead of conducting such a complete analysis, users are more apt to use their intuitive understanding of the resource and each other to experiment with different rule changes until they find a combination that seems to work in their setting.

To understand the types of tools that are available to users somewhat better, let us examine in some detail the kind of boundary, authority, payoff, and position rules used in field settings. These four clusters of rules are the major tools used to affect the management of common-pool resources, while information, scope, and aggregation rules are utilized to complement changes induced by these four rules.

For the past 14 years, colleagues at or associated with the Workshop in Political Theory and Policy Analysis at Indiana University have studied a very large number of fisheries, forests, irrigation systems, and groundwater basins, as well as other common-pool resources. We have collected an immense archive of original case studies written by many scholars on all sectors in all parts of the world (Hess, 1999). We developed structured coding forms to help us identify the specific kinds of common-pool situations faced in the field, as well as the types of rules that users have evolved over time to try to govern and manage these resources effectively. In order to develop standardized coding forms, we read hundreds of cases describing how local common-pool resources were or were not regulated by a government

---

[2] This list of rules is the result of many years of theoretical and empirical work on a framework for conducting institutional analyses (see E. Ostrom *et al.*, 1994: Chap. 2).

agency, by the users themselves, or by a nongovernmental organization (NGO).

A. Affecting the Characteristics of Users through Boundary Rules

The most frequent recommendation concerning boundary rules in the policy literature is to limit the number of persons allowed to appropriate from a common-pool resource so that the level of appropriation is reduced or to require users to obtain a license before harvesting. Boundary rules affect the types of participants with whom other participants will be interacting. If contingent cooperation is perceived to be a possibility, then one of the most important ways to enhance the likelihood of using reciprocity norms is to increase the proportion of participants who are well known in a community, have a long-term stake in that community, and would find it costly to have their reputation for trustworthiness harmed in that community. Reducing the number of users while opening the resource to strangers who are willing to pay a license fee, but who lack a long-term interest in the sustainability of a particular resource, may reduce the level of trust and willingness to use reciprocity and thus increase enforcement costs substantially.

We have identified 27 boundary rules as having been used in at least one common-pool resource somewhere in the world (E. Ostrom *et al.*, 1989). While some systems use only a single boundary rule, many use two or three of these rules in combination. Boundary rules can be broadly classified in three general groups defining how individuals gain authority to enter and appropriate resource units from a common-pool resource. The first type of boundary rule relates to an individual's citizenship, residency, or membership in a particular organization. Many forestry and fishing user groups require members to have been born in a particular location. A second broad group of rules relates to individual ascribed or acquired personal characteristics. Other user groups may require that appropriation depends on ethnicity, clan, or caste. A third group of boundary rules relates to the relationship of an individual with the resource itself. Using a particular technology or acquiring appropriation rights through an auction or a lottery are examples of this type of rule. About half of the rules relate to the characteristics of the users themselves. The other half involve diverse relationships with the resource.

In a systematic coding of those case studies for which sufficient information existed about rules related to inshore fisheries in many parts of the world, E. Schlager coded 33 user groups out of the 44 groups identified as having at least some rules regarding the use of the resource. All 33 groups depended on some combination of 14 different boundary rules (Schlager, 1994: 258). None of these groups relied on a single boundary rule. Thirty out of 33 groups (91%) limited fishing to those individuals who lived in a nearby community, while 13 groups also required membership in a local organization. Consequently, most inshore fisheries organized by the users themselves restrict fishing to those individuals who are well known to each other, have a relatively long-term time horizon, and are connected to one another in multiple ways.

After residency, the next most frequent type of rule, used in two-thirds of the organized subgroups, involves the type of technology that a potential fisher must use. These rules are often criticized by policy analysts, since gear restrictions tend to reduce the "efficiency" of fishing. Gear restrictions have many consequences, however. Used in combination with authority rules that assign fishers using one type of gear to one area of the fishing grounds and fishers using another type of gear to a second area, they solve conflicts among noncompatible technologies. Many gear restrictions also place a reduced load on the fishery itself and thus help to sustain longer-term use of the resource.

Other rules were also used. A scattering of groups used ascribed characteristics (age, two groups; ethnicity, three groups; race, five groups). Three types of temporary use rights included government licenses (three groups), lottery (five groups), and registration (four groups). Seven groups required participants to have purchased an asset such as a fishing berth, while three groups required ownership of nearby land. Schlager did not find that any particular boundary rule was correlated with higher performance levels across the groups. She did find, however, that the 33 groups who had at least one boundary rule tended to be able to solve common-pool problems more effectively than the 11 groups who had not crafted boundary rules.

Thus, many of the rich diversity of boundary rules used by users in the field attempt to ensure that the users will be relating to others who live nearby and have a long-term interest in sustaining the productivity of the resource. One way of coping with the commons is thus changing the composition of who uses a common-pool resource to increase the proportion of participants who have a long-term interest, who are more likely to use reciprocity, and who can be trusted. Central governments tend to use a smaller set of rules and some of these may open up a resource to strangers without a longer-term commitment to the resource.

B. Affecting the Set of Allowable Actions through Authority Rules

Authority rules are also a major type of rule used to regulate common-pool resources. We identified a diversity of authority rules used in field settings. Some rules involve a simple formula. Many forest resources, for example, are closed to all forms of harvesting during one portion of the year and open for extraction by all who meet the boundary rules during an open season. Most authority rules, however, have two components: an assignment and a basis. A fisher, for example, might be assigned to a fixed location (a fishing spot) or to a fixed rotational schedule, or a member of the founding clan may be assigned (allowed) to cut timber anywhere in a forest. In addition to the assignment used in an authority rule, most rules required a basis for the assignment. For example, a fisher might be assigned to a fixed location based on a number drawn in a lottery, on the purchase of that spot in an auction, or on the basis of his or her historical use.

If all of the bases were likely to be combined with all of the assignment possibilities, there would be 112 different authority rules (8 assignment formulas times 14 bases). A further complication is that the rules for one product may differ from those of another product in the same resource. In regard to forest resources, for example, children may be authorized to pick fruit from any tree located in a forest so long as it is for their own consumption, women may be authorized to collect so many headloads of dead wood for domestic firewood and certain plants for making crafts, while *shaman* are the only ones authorized to collect medicinal plants from a particular location in a forest. Appropriation rights to fish are frequently related to a specific species. Thus, the exact number of rules that are actually used in the field is difficult to compute since not all bases are used with all formulas, but many rules focus on specific products. A still further complication is that the rules may regularly change over the course of a year, depending on resource conditions.

Schlager (1994: 259–260) found that all 33 organized subgroups used one of five basic assignments in their authority rules. Every user group included in her study assigned fishers to fixed locations using a diversity of bases, including technology, lottery, or historical use. Thus, spatial demarcations are a critical variable for inshore fisheries. Nine user groups required fishers to limit their harvest to fish that met a specific size requirement, while seven groups allocated fishers to fishing spots using a rotation system and seven other groups only allowed fishing locations to be used during a specific season. Four groups allocated fishing spots for a particular time period (a fishing day or a fishing season).

An important finding—given the puzzles addressed in this article—is that the authority rule most frequently recommended by policy analysts is *not* used in any of the coastal fisheries included in Schlager's study. No attempt was made "by the fishers involved to directly regulate the quantity of fish harvested based on an estimate of the yield. This is particularly surprising given that the most frequently recommended policy prescription made by fishery economists is the use of individual transferable quotas based on estimates on the economically optimal quantity of fish to be harvested over the long run" (Schlager, 1994: 265). In an independent study of 30 traditional fishery societies, James Acheson and colleagues also noted the surprising absence of quota rules:

> All of the rules and practices we found in these 30 societies regulate "how" fishing is done. That is, they limit the times fish may be caught, the locations where fishing is allowed, the technology permitted, and the stage of the life cycle during which fish may be taken. None of these societies limits the "amount" of various species that can be caught. Quotas—the single most important concept and tools of scientific management—is conspicuous by its absence (Acheson *et al.*, 1998: 397).

Local inshore fishers, when allowed to manage a riparian area, thus use rules that differ substantially from those recommended by advocates of scientific management. Fishers have to know a great deal about the ecology of their inshore region, including spawning areas, nursery areas, the migration routes of different species, and seasonable patterns, just to succeed as fishers. Over time, they learn how "to maintain these critical life-cycle processes with rules controlling technology, fishing locations, and fishing times. Such rules in their view are based on biological reality" (Acheson *et al.*, 1998: 405).

The diversity of rules devised by users greatly exceeds the limited authority rules recommended in textbook treatments of this problem. Users thus cope with the commons by a wide variety of rules affecting the actions available to participants and thus their basic set of strategies. Given this wide diversity of rules, it is particularly noteworthy that rules assigning users a right to a specific quantity of a resource are used so infrequently in inshore fisheries. (They are used more frequently when allocating forest products, where the quantity available, as well as the quantity harvested, is much easier to measure.) To assign a user a specific quantity of a resource unit requires that those making the assignment know the total available units.

C. Affecting Outcomes through Payoff and Position Rules

One way to reduce or redirect the appropriations made from a common-pool resource is to change payoff rules so as to add a penalty to actions that are prohibited. Many user groups also adopt norms that those who are rule breakers should be socially ostracized or shunned and individual users tend to monitor each other's behavior rather intensively. Three broad types of payoff rules are used extensively in the field: (1) the imposition of a fine, (2) the loss of appropriation rights, and (3) incarceration. The severity of each of these types of sanctions can range from very low to very high and tends to start out on the low end of the scale. Inshore fisheries studied by Schlager relied heavily on shunning and other social norms and less on formal sanctions.

Passing rules that impose costs is relatively simple. The real difficult task is monitoring behavior to ascertain if rules are being broken. Self-organized fisheries tend to rely on self-monitoring more than the creation of a formal position of guard. Most inshore fishers now use shortwave radios as a routine part of their day-to-day operations, allowing a form of instant monitoring to occur. An official of a West Coast Indian tribe reports, for example, that "it is not uncommon to hear messages such as 'Did you see so-and-so flying all that net?' over the shortwave frequency—a clear reference to a violation of specified gear limits" (cited in Singleton, 1998: 134). Given that most fishers will be listening to their shortwave radio, "such publicity is tantamount to creating a flashing neon sign over the boat of the offender. Such treatment might be proceeded or followed by a direct approach to the rule violator, advising him to resolve the problem. In some tribes, a group of fishermen might delegate themselves to speak to the person" (cited in Singleton, 1998: 134).

Among self-organizing forest governance systems, creating and supporting a position as guard is frequently essential, since resource units are highly valuable and a few hours of stealth can generate substantial illicit income. Monitoring rule conformance among forest users by officially designated and paid guards may make the difference between a resource in good condition and one that has become degraded. In a study of 279 forest *panchayats* in the Kumaon region of India, Agrawal and Yadama (1997) found that the number of months a guard was hired was the most important variable affecting forest conditions. The other variables that affected forest conditions included the number of meetings held by the forest council (a time when infractions are discussed) and the number of residents in the village.

> It is evident from the analysis that the capacity of a forest council to monitor and impose sanctions on rule-breakers is paramount to maintaining the forest in good condition. Nor should the presence of a guard be taken simply as a formal mechanism that ensures greater protection. It is also an indication of the informal commitment of the *panchayat* and the village community to protect their forests. Hiring a guard costs money.... If there was scant interest in protecting the forest, villagers would have little interest in setting aside the money necessary to hire a guard (Agrawal and Yadama, 1997: 455).

Boundary and authority rules also affect how easy or difficult it is to monitor activities and impose sanctions on rule infractions. Closing a forest or an inshore fishery for a substantial amount of time, for example, has multiple impacts. It protects particular plants or fish during critical growing periods and allows the entire system time to regenerate without disturbance. Further, during the closed season, rule infractions are highly obvious to anyone as any user in the resource is almost certainly breaking the rules. Similarly, requiring users to use a particular technology may reduce the pressure on the resource, help to solve conflicts among users of incompatible technologies, and also make it very easy to ascertain if rules are being followed. Changing payoff rules is the most direct way of coping with commons dilemmas.

D. Affecting Outcomes through Changes in Information, Scope, and Aggregation Rules

Information, scope, and aggregation rules tend to be used in ways that complement changes in boundary, authority, payoff, and position rules. Individual systems vary radically in regard to the mandatory information that they require. Many smaller and informal systems rely entirely on a voluntary exchange of information and on mutual monitoring. Where resource units are valuable and the size of the group is large, more and more requirements are added regarding the information that must be kept by users or their officials. Scope rules are used to limit harvesting activities in some regions that are being treated as refugia. By not allowing any appropriation from these locations, the regenerative capacity of a system can be enhanced. Aggregation rules are used extensively in collective-choice processes and less extensively in operational settings, but one aggregation rule that is found in diverse systems is a requirement that harvesting activities be done in teams. This increases the opportunity for mutual monitoring and reduces the need to hire special guards.

It is important to note that repeated studies have not yet found any *particular* rules to have a statistically positive relationship to performance. The absence of any boundary or any authority rules is consistently associated with poor performance. Relying on only a single type of rule for an entire set of common-pool resources is also negatively related.

IV. THE DAUNTING SEARCH FOR BETTER RULES

It should now be obvious that the search for rules that improve the outcomes obtained in commons dilemmas is an incredibly complex task involving a potentially infinite combination of specific rules that could be adopted. To ascertain whether one has found an optimal set of rules to improve the outcomes achieved in a single situation, one would need to analyze how diverse rules affect incentives, strategies, and outcomes. Since there are multiple rules that affect incentives, conducting such an analysis would be an incredibly time- and resource-consuming process. For example, if only five changes in rules per component were considered, there would be 5^7 or 75,525 different rule configurations to analyze. This is a gross simplification, however, since some of the important rules used in field settings include more than 25 rules (in the case of boundary rules) and even over 100 variants (in the case of authority rules). Further, how these changes affect the outcomes achieved in a particular location depends on the biophysical characteristics of that location and the type of community relationships that already exist. No set of policy analysts (or even all of the policy analysts in the world today) could ever have sufficient time or resources to analyze over 75,000 combinations of rule changes and resulting situations, let alone all of the variance in these situations due to biophysical differences.

V. EXPERIMENTING WITH RULE CHANGES

Instead of assuming that designing rules that approach optimality, or that even improve performance, is a relatively simple analytical task that can be undertaken by distant, objective analysts, we need to understand the policy design process as involving an effort to tinker with a large number of component parts (see Jacob, 1977). Those who tinker with any tools—including rules—are trying to find combinations that work together more effectively than other combinations. Policy changes are experiments based on more or less informed expectations about potential outcomes and the distribution of these outcomes for participants across time and space (Campbell, 1969, 1975). Whenever individuals agree to add a rule, change a rule, or adopt someone else's proposed rule set, they are conducting a policy experiment. Further, the complexity of the ever-changing biophysical world combined with the complexity of rule systems means that any proposed rule change faces a nontrivial probability of error.

When there is only a single governing authority, policymakers have to experiment simultaneously with *all* of the common-pool resources within their jurisdiction with each policy change. And, once a change has been made and implemented, further changes will not be made rapidly. The process of experimentation will usually be slow, and information about results may be contradictory and difficult to interpret. Thus, an experiment that is based on erroneous data about one key structural variable or one false assumption about how actors will react can lead to a very large disaster (see Wilson *et al.*, 1999). In any design process where there is substantial probability of error, having redundant teams of designers has repeatedly been shown to have considerable advantage (see Landau, 1969, 1973; Bendor, 1985).

VI. SELF-ORGANIZED RESOURCE GOVERNANCE SYSTEMS AS COMPLEX ADAPTIVE SYSTEMS

As discussed in Section II, the very concept of *organization* is closely tied for many scholars to the presence of a central director who has designed a system to operate in a particular way. Consequently, the mechanisms used by noncentrally directed systems are not always well understood. Many self-organized resource governance systems are invisible to the officials of their own country or those from donor agencies. A classic example of this occurred in the Chitwan valley of Nepal several years ago when an Asian Development Bank team of irrigation engineers recommended a very large loan to build a dam across the Rapti River to enable the farmers there to irrigate their crops. What the engineering design team did not see was 85 farmer-managed irrigation systems that already existed in the valley and that had achieved relatively high performance. Most farmers in the Chitwan valley already obtained three irrigated crops a year as a result of their participation in the activities of these irrigation systems (see Benjamin *et al.*, 1994).

In contrast to forms of organization that are the result of central direction, most self-organized groups—including the types of locally organized users of fisheries and forests discussed in this article—are better viewed as complex adaptive systems. These systems are composed of a large number of active elements whose rich patterns of interactions produce emergent properties that are not easy to predict by analyzing the separate parts of a system. Holland (1995: 10) views complex adaptive systems as "systems composed of interacting agents described in terms of rules. These agents adapt by changing their rules as experience accumulates." Such systems "exhibit coherence under change, via conditional action and anticipation, and they do so without central direction" (Holland, 1995: 38–39). Holland points out that complex adaptive systems differ from physical systems that are not adaptive and that have been the foci of most scientific effort. It is the physical sciences that have been the model for many aspects of contemporary social science. Thus, the concepts needed to understand the adaptivity of systems are not yet well developed by social scientists.

A. Properties and Mechanisms of Complex Adaptive Systems

No general theory of complex adaptive systems yet exists to provide a coherent explanation for processes shared by all complex adaptive systems. Biologists have studied many different adaptive systems but within separate fields of biology. Thus, even biologists have not recognized some of the similarities of structures and processes that characterize the central nervous system, the immune system, and the evolution of species. Recent work at the Sante Fe Institute has begun to identify central attributes, mechanisms, and processes used by all complex adaptive systems including both biological systems, as well as markets and other social systems that are not centrally directed.

It appears that all complex adaptive systems share four basic properties: nonlinearity, flows, diversity, and aggregation. The first three properties are self-evident and clearly characterize the types of self-organized resource governance systems discussed in this article. *Aggregation* refers to the "emergence of complex larger-scale behavior from the aggregate interactions of less complex agents" (Holland, 1995: 11). For example, many irrigation systems are divided into several tiers and multiple units at each of these tiers. All of the farmers on a field irrigation channel are responsible for distributing the water to this small channel as well as keeping it in good repair. All farmers whose field channels are served by a branch canal may send a representative to a branch canal organization that focuses its attention on the distribution of water among all branches and on the maintenance of the distribution canals. The branch canal organization may send a representative to a central committee who is responsible for the headworks that divert the water from a river into the system in the first place. The rules used on one branch canal or one field channel may be quite different than on others. There is no single center of authority for these systems that makes all relevant decisions on how to get water from the river to a farmer's field, but in many farmer-organized systems the water is distributed in an organized fashion and all of the waterworks are maintained as a result of the aggregation of decisions and actions at multiple levels.

In addition to these four attributes, complex adaptive systems also use three mechanisms that are key to the adaptive process itself. These include the use of tags, internal models, and building blocks.

1. The Use of Tags

Tagging is a universal mechanism for boundary formation and aggregation of units in complex adaptive systems. "Tags are a pervasive feature of [complex adaptive systems] because they facilitate selective interactions. They allow agents to select among agents or objects that would otherwise be indistinguishable" (Holland, 1995: 14). All of the types of boundary rules discussed here involve the specification of the tags that will be used to determine who is authorized to be a co-user from a common-pool resource. Residency, prior membership, and personal characteristics are attributes that already exist and are easy to use as boundary tags. Local governance systems rely heavily on tags that identify individuals who are already known to each other, who have a long-term stake in the sustainability of a resource, who have an incentive to build a reputation for being trustworthy, and who are thus likely to extend reciprocity rather than recalcitrance in dealing with joint problems.

Tags are also used extensively to mark locations in a resource, to warn rule infractors, and even to mark individual organisms that need to be treated in a special way. An example of the latter occurs along the Maine coast, where it is forbidden to harvest berried lobsters (those with eggs). Such lobsters are V-notched and returned back to the sea. Any other fisher who captures a V-notched lobster is also supposed to return it to the sea.

2. Internal Models

The users from a common-pool resource build internal models of the resource, of the relationships among the components of the resource, and of where their own actions positively or negatively affect one another and the resource. Among the shared lore for most fishing villages is a clear understanding of where fish breed, where young fish tend to cluster, the length of time it takes for fish to be mature and reproduce, the migration patterns of fish, the food chain in a location, and other information. Many inshore fishers develop their own maps of all of the fishing spots in their grounds. In an effort to reduce the interference of one boat with another boat's fishing, these are frequently defined so that if all "spots" are filled, all boats are still able to have a good chance to catch fish. These maps are then used in a variety of allocation rules that specify the basis for how any particular boat is assigned to a particular fishing spot. Users of forests also map their forests and may create refugia—sometimes as sacred forests—for sections of a forest that are particularly rich in biodiversity. By not harvesting from these refugia, they serve as a source of regeneration to other nearby locations that are disturbed through harvesting.

3. Building Blocks

Building blocks are ways of breaking down complex processes into small chunks that can be used in multiple ways and combined and re-combined repeatedly and at diverse levels. Once an authority rule that allocates resource units on some basis is determined, for example, using the same basis again to allocate responsibility for maintenance work is considered to be a fair allocation of benefits and costs in many cultures and is relatively easy to remember. On the large Chhattis Mauja farmer-organized system in Nepal, for example, water was originally allocated by the land area served. In the 1950s, the formula used for maintenance work was that each branch canal was responsible to send one person to work on the main canal for each 17 hectares of area it irrigated. "The term used for a person-day of labor for canal maintenance was *kulara*. Since the share of water a branch canal is entitled to receive is the same as the resource mobilization requirement, water allocation is now also referred to as 'so many *kulara* of water' " (Yoder, 1991: 7). As the system has grown, the total number of *kularas* has now been set at 177 shared among 44 branch canals. Voting rights are now also set in terms of *kularas*. "Therefore, a branch canal with five *kulara* was entitled to 5/177 of the water in the main canal, responsible to supply 5/177 of the resources mobilized for the irrigation, and had five of the total 177 votes in all important decisions" (Yoder, 1991: 7).

B. Changing Rules as an Adaptive Process

Given the logic of combinatorics, it is impossible—as we showed earlier—to conduct a complete analysis of the expected performance of all of the potential rule changes that could be made by the individuals who are served by a self-organized resource governance system and are trying to improve its performance. A similar impossibility also exists for many biological systems. Let us explore these similarities.

Self-organizing resource governance systems have two structures that are somewhat parallel in function to the concepts of a genotype and a phenotype in biology. Phenotypic structures characterize an expressed organism—how bones, organs, and muscles develop, relate, and function in an organism in a particular environment. The components of an action situation characterize an expressed situation—how the number of participants, the information available, and their opportunities and costs create incentives, and how incentives lead to types of outcomes in a particular environment. The genotypic structure characterizes the set of instructions encoded in DNA to produce an organism with a particular phenotypic structure. A rule configuration is a set of instructions on how to produce the structure of relationships among individuals in an action situation, which is also affected by the biophysical world and the kind of community or culture in which an action situation is located.

The evolution of social systems does not follow the same mechanisms as the evolution of species. As an evolutionary process, of course, there must be the generation of new alternatives, selection among new and old combinations of structural attributes, and retention of those combinations of attributes that are successful in a particular environment. In evolving biological systems, genotypic structures are changed through mechanisms such as crossover and mutation, and the distribution of particular types of instructions depends on the survival rate of the phenotypes that they produce in given environments. Instead of blind variation, however, human agents do try to use reason and persuasion in their efforts to try to devise better rules, but the process of choice always involves experimentation.

Rule configurations can be represented as a string of symbols that describe which rules are in effect in a particular location. In Table I, the left-hand column lists a small subset of the rules typically found in the field. A "1" has been entered in a cell of a column to indicate when a specific rule is in use, and a "0" has been entered when the rule is not in use. Thus, the eight columns represent the simplified rule configurations for eight hypothetical locations.

Location A would be characterized as one at which the participants are required both to live in a local community and to use geographic assignments or a rotation system, but also at which there are no formal sanctions used or guards hired to monitor conformance to the entry rules and rotation system. Further, no records are kept, there are no refugia, and users can work individually or in teams. In contrast, Location C creates a refugia in which no appropriation may be undertaken, and does not allocate the remaining space to individual users. If there is a very good breeding and nursery area in this fishery, and it can be well demarcated so that it is obvious if someone has ventured into the refugia by error (or by intent), a rule system that does not limit fishing in any other way may be a very effective system for protecting the regenerative capabilities of a fishery. It is certainly less expensive than naming specific fishing spots, making an allocation of each fishing team to particular spots, and enforcing a more complicated rule system. At Location C, users also need to live in the local area, and they do face fines if a fellow user notices that they have appropriated from the refugia. Though a guard is not created, records are kept of any infractions noted through the mutual monitoring process.

TABLE I

Rule Configurations and Appropriation Situations[a]

Rule configurations	Locations A	B	C	D	E	F	G	H
Boundary rules								
Local residency	1	1	1	1	0	0	0	0
Member of local co-op	0	1	0	1	1	1	1	1
Authority rules								
Spatial assignment	1	0	0	0	1	0	1	0
Rotate in time	1	1	0	0	1	1	0	1
Payoff rules								
Fines	0	0	1	1	0	0	1	1
Loss of rights	0	0	0	0	0	0	1	1
Position rules								
All symmetric	1	1	1	1	0	0	0	0
Guard created	0	0	0	0	1	1	1	1
Information rules								
Infractions are recorded	0	0	0	1	1	0	0	1
Appropriation is recorded	0	0	0	0	0	0	0	1
Scope rules								
Forbidden access to refugia	0	0	1	0	1	0	1	1
Aggregation rules								
Must appropriate in teams	0	0	0	0	0	0	0	1

[a] A "1" indicates when a specific rule is in use, and a "0" indicates when the rule is not in use. See text for details.

Location H utilizes a much fuller array of rules. Users must belong to a local cooperative, even though they do not have to live in the local community. Both a refugia and a system for allocating time in which to conduct appropriation have been created, although, with the exception of the forbidden area of the refugia, the latter is not spatially defined. Payoff rules have been modified so that there are both fines for infractions as well as the possibility of losing rights to appropriate, if infractions are substantial. Records of infractions and of the time when someone appropriates are required, and individuals are constrained to working in teams rather than being allowed to appropriate as individuals.

Whether any or all of these rule configurations structured the incentives of users so that participants appropriate from their resource in a sustainable and efficient manner cannot be ascertained from learning about the rules alone. Whether a set of rules enhances performance depends on the structure of the biophysical system in that location and the willingness of participants to follow the rules. Whether the strategy of creating a refugia, as in C and H, works better than the geographic assignment, as in A, depends on how large the refugia is, how rapidly and extensively regeneration from the refugia spreads through the rest of the resource, and how well users understand each of these systems and try to make them work. All three rule systems are likely to work well in some locations and poorly in others.

With this form of representation we can begin to see how experimentation with rules may be similar to structural changes affecting other types of complex adaptive systems. Most systems are likely to start with one or two very simple rules. An obvious first candidate is to use tags to close the boundary to outsiders so that the likelihood of contingent cooperation and conformance to agreements will be enhanced. By only changing a few rules at the beginning, everyone can come to understand those rules while they are evaluating how they work. A second obvious candidate is to use the shared model of the environment built up through years of interaction in an environment to refine where harvesting should be undertaken and when. Space and time are obvious candidates for allocating access to resources in a manner that is relatively low in cost to sustain. If the community is small enough and shares common norms at a high enough level, creating formal sanctions, guards, records, and other rules may not be necessary. Thus, one can imagine a process in which a rule configuration of mostly zeros slowly converts over time to a scattering of ones and potentially to a rule configuration with many ones.

Changes in specific rules may come about through accident (such as forgetting a rule) or through specific collective-choice processes in which considerable time and effort are devoted to considering why performance needs to be enhanced and which rules might be changed. Since many users will have experience with more than one product (e.g., mushrooms from a forest), rules tested with regard to one product may also be applied with regard to others, if they are successful. Migration of individuals into a community brings individuals with repertoires of different rules used in other locations. Commerce with other groups lets users see and learn about others who may be doing better (or worse) than they in regulating a sustainable biodiverse and efficient resource system. Thus, a self-organized resource governance system with a higher level of inmigration or greater communication with other localities is more likely to adapt and change rules over time than one in which few new ideas concerning how to use rules as tools are brought into the system. Trial-and-error processes may give relatively rapid feedback about rules that obviously do not work in a particular environment, but this is not always the case when the effect of human action on the environment has a long time delay. If all self-organized resource governance systems are totally independent and there is no communication among them, then each has to learn through its own trial-and-error process. Many will find that rules that they have tried do not work. Some will fail entirely.

The rate of change will differ among self-organized resource governance systems. As with all learning, the rate of change is an important variable that affects performance over time. If change occurs too rapidly, little is learned from each experiment before another experiment is launched. Respect for tradition and even religious mystification has been used to increase the retention of rules that are considered by at least some participants to work better. If the heavy hand of tradition is, however, too heavy and squelches innovation, a system that may have been well adapted to a past environment may find itself faltering as external changes occur without internal changes also occurring.

VII. THE ADVANTAGES AND LIMITS OF PARALLEL SETS OF LOCAL USERS IN POLICY EXPERIMENTS

Let us now discuss why a series of relatively autonomous, self-organized, resource governance systems may do a better job in policy experimentation than a single central authority. Among the advantages of authorizing the users of smaller-scale common-pool resources to adopt policies that regulate the use of these resources are:

- Local knowledge. Users who have lived and appropriated from a resource system over a long period of time have developed relatively accurate mental models of how the biophysical system itself operates, since the very success of their appropriation efforts depends on such knowledge. They also know others living in the area well and what norms of behavior are considered appropriate.
- Inclusion of trustworthy participants. Users can devise rules that increase the probability that others are trustworthy and will use reciprocity. This lowers the cost of relying entirely on formal sanctions and paying for extensive guarding.
- Reliance on disaggregated knowledge. Feedback about how the resource system responds to changes in actions of users is provided in a disaggregated way. Fishers are quite aware, for example, if the size and species distribution of their catch is changing over time.
- Better-adapted rules. Given the preceding, users are more likely to craft rules that are better adapted to each of the local common-pool resources than any general system of rules.
- Lower enforcement costs. Since local users have to bear the cost of monitoring, they are apt to craft rules that make infractions obvious so that monitoring costs are less. Further, by creating rules that are seen as legitimate, rule conformance will tend to be higher.
- Redundancy. The probability of failure throughout

an entire region is greatly reduced by the establishment of parallel systems of rule making, interpretation, and enforcement.

There are, of course, limits to all ways of organizing the governance of common-pool resources. Among the limits of a highly decentralized system are:

- Some users will not organize. While the evidence from the field is that many local users do invest considerable time and energy into their own regulatory efforts, other groups of users do not do so. There appear to be many reasons why some groups do not organize, including the presence of low-cost alternative sources of income and thus a reduced dependency on the resource, considerable conflict among users along multiple dimensions, lack of leadership, and fear of having their efforts overturned by outside authorities.
- Some self-organized efforts will fail. Given the complexity of the task involved in designing rules, some groups will select combinations of rules that generate failure instead of success. They may be unable to adapt rapidly enough to avoid the collapse of a resource system.
- Local tyrannies. Not all self-organized resource governance systems will be organized democratically or rely on the input of most users. Some will be dominated by a local leader or a power elite who only change rules that they think will give them further advantage. This problem is accentuated in locations where the cost of exit is particularly high and is reduced where users can leave when local decision makers are not responsible to a wide set of interests.
- Stagnation. Where local ecological systems are characterized by considerable variance, experimentation can produce severe and unexpected results, leading users to cling to systems that have worked relatively well in the past and to stop innovating long before they have developed rules likely to lead to better outcomes.
- Inappropriate discrimination. The use of identity tags is frequently an essential method for increasing the level of trust and rule conformance. However, tags based on ascribed characteristics can be the basis for excluding some individuals from access to sources of productive endeavor that has nothing to do with their trustworthiness.
- Limited access to scientific information. Whereas time and place information may be extensively developed and used, local groups may not have access to scientific knowledge concerning the type of resource system involved.
- Conflict among users. Without access to an external set of conflict-resolution mechanisms, conflict within and across common-pool resource systems can escalate and provoke physical violence. Two or more groups may claim the same territory and may continue to make raids on one another over a very long period of time.
- Inability to cope with larger-scale common-pool resources. Without access to some larger-scale jurisdiction, local users may have substantial difficulties regulating only a part of a larger-scale common-pool resource. They may not be able to exclude others who refuse to abide by the rules that a local group would prefer to use. In this situation, local users have no incentives to restrict their own use and watch others take away all of the valued resource units that they have not appropriated.

VIII. THE CAPABILITIES OF POLYCENTRIC SYSTEMS IN COPING WITH TRAGEDIES OF THE COMMONS

Many of the capabilities of a parallel adaptive system can be retained in a polycentric governance system. A polycentric system is one in which citizens are able to organize not just one but multiple governing authorities at differing scales (see McGinnis, 1999; V. Ostrom, 1997). Each unit may exercise considerable independence to make and enforce rules within a circumscribed scope of authority for a specified geographical area. In a polycentric system, some units are general-purpose governments and others may be highly specialized. Self-organized resource governance systems, in such a system, may be special districts, private associations, or parts of a local government. These are nested in several levels of general-purpose governments that also provide civil, equity, and criminal courts.

In a polycentric system, the users of each common-pool resource would have some authority to make at least some of the rules that are related to how that particular resource will be utilized, and thus would achieve most of the advantages of utilizing local knowledge, and the redundancy and rapidity of a trial-and-error learning process. On the other hand, problems associated with local tyrannies and inappropriate discrimination can be addressed in larger, general-purpose governmental units that are responsible for protecting the rights of all citizens and for the oversight of appropriate exercises of authority within smaller units of government. It is also possible to make a more effective blend of scientific information with local knowledge where major universities and research stations are located in larger units but have a responsibility to relate recent scientific findings to multiple smaller units within their region. Because polycentric systems have overlapping units, information about what has worked well in one setting can be transmitted to others who may try it out in their settings. Associations of local resource governance units can be encouraged to speed up the exchange of information about relevant local conditions and about policy experiments that have proved particularly successful. And, when small systems fail, there are larger systems to call upon—and vice versa.

Polycentric systems are themselves complex adaptive systems without one central authority dominating all of the others. Thus, there is no guarantee that such systems will find the optimal combination of rules at diverse levels that are optimal for any particular environment. In fact, one should expect that all governance systems will be operating at less-than-optimal levels, given the immense difficulty of fine-tuning any very complex, multitiered system.

In the United States, there are many examples of dynamic, polycentric resource governance systems that display strong evidence of high performance. One example is the Maine lobster fishery, which is noteworthy because of the long-term, complementary roles adopted by both local and state governance systems. Maine is organized into riparian territories along most of the coast. Boundary rules and many of the day-to-day fishing regulations are organized by harbor gangs.

> In order to go fishing at all, one must become a member of a "harbor gang," the group of fishermen who go lobstering from a single harbor. Once one has gained admittance into such a group, one can only set traps in the traditional territory of that particular harbor gang. Members of harbor gangs are expected to obey the rules of their gang concerning fishing practices, which vary somewhat from one part of the coast to another.... There is strong statistical evidence that the territorial system, which operates to limit the number of fishers exploiting lobsters in each territory, helps to conserve the lobster resource (Acheson *et al.*, 1998: 400).

At the same time, the State of Maine has long established formal laws that protect the breeding stock and increase the likelihood that regeneration rates will be high. "At present, the most important conservation laws are minimum and maximum size measures, a prohibition against catching lobsters with eggs, and a law to prohibit the taking of lobsters which once had eggs and were marked—i.e. the 'V-notch' law" (Acheson *et al.*, 1998: 400). Neither the state nor any of the harbor gangs has tried to limit the quantity of lobster captured. The state does not make any effort to limit the number of fishers since this is already done at a local level. However, the state has been willing to intercede when issues exceed the scope of control of local gangs. In the late 1920s, for example, when lobster stocks were at very low levels and many local areas appear to have had substantial compliance problems, the state took a number of steps—including threats to close the fishery—that supported informal local enforcement efforts. By the late 1930s, compliance problems were largely resolved and stocks had rebounded (although it cannot be shown that these two results are related, just correlated).

Recently, in response to changes that were breaking down the harbor gang system, the state formalized the system by dividing the state into zones with democratically elected councils. Each council was given authority over rules that have principally local impacts—trap limits, days and times fished, and so on. Interestingly, the formalization of local zones was followed, almost immediately, by the creation of an informal council of councils to address problems at a greater than local scale. It is expected that this council of councils will be formalized soon.

Today the state uses only about six patrol officers on the water to police the activities of 6800 lobstermen, all the other fisheries, and coastal environmental laws. During the 1990s, the lobster fishery has been growing substantially with increased yields. At the same time, there is strong evidence that the number of reproductive age females in Maine waters is very large and that the recruitment will continue at a high level.

IX. CONCLUSIONS

Widespread concern for conserving biodiversity is a reflection of the inherent and ecological value of such diversity. Those who advocate the necessity of sustaining biodiversity have frequently called for policy reforms that reduce the diversity of institutional arrangements for governing and managing complex common-pool resources. Yet, as Ashby long ago established, regulators need as much variety in their response capabilities as the systems they are regulating. To achieve biodiverse resource systems, we need to protect and enhance the institutional diversity of evolved governance regimes within the context of a broader multitiered, polycentric governance regime.

See Also the Following Articles

COMMONS, THEORY AND CONCEPT OF • GOVERNMENT LEGISLATION AND REGULATION • MARKET ECONOMY AND BIODIVERSITY

Acknowledgments

Support for the research on which this article is based has been provided by the MacArthur Foundation and the Latin American and Caribbean Regional Office of the Food and Agriculture Organization (FAO) of the United Nations. My thanks to Patty Dalecki and Tom Hargis-Young for editing and preparing this manuscript.

Bibliography

Acheson, J. M., Wilson, J. A., and Steneck, R. S. (1998). Managing chaotic fisheries. In *Linking Social and Ecological Systems. Management Practices and Social Mechanisms for Building Resilience* (F. Berkes and C. Folke, eds.), pp. 390–413. Cambridge University Press, Cambridge, United Kingdom.

Agrawal, A. (1994). Rules, rule making, and rule breaking: Examining the fit between rule systems and resource use. In *Rules, Games, and Common-Pool Resources* (E. Ostrom, R. Gardner, and J. M. Walker, eds.), pp. 267–282. University of Michigan Press, Ann Arbor.

Agrawal, A., and Yadama, G. N. (1997). How do local institutions mediate market and population pressures on resources? Forest *panchayats* in Kumaon, India. *Development and Change* 28(3), 435–465.

Arnold, J. E. M. (1998). *Managing Forests as Common Property*. FAO Forestry Paper No. 136. Food and Agriculture Organization of the United Nations, Rome.

Ashby, W. R. (1960). *Design for a Brain: The Origin of Adaptive Behavior*. John Wiley & Sons, New York.

Bendor, J. B. (1985). *Parallel Systems. Redundancy in Government*. University of California Press, Berkeley.

Benjamin, P., Lam, W. F., Ostrom, E., and Shivakoti, G. (1994). *Institutions, Incentives, and Irrigation in Nepal. Decentralization: Finance & Management Project Report*. Associates in Rural Development, Burlington, Vermont.

Bromley, D. W., Feeny, D., McKean, M. A., Peters, P., Gilles, J., Oakerson, R. J., Runge, C. F., and Thomson, J. T. (eds.). (1992). *Making the Commons Work: Theory, Practice, and Policy*. ICS Press, San Francisco.

Campbell, D. T. (1969). Reforms as experiments. *Am. Psychologist* 24(4), 409–429.

Campbell, D. T. (1975). On the conflicts between biological and social evolution and between psychology and moral tradition. *Am. Psychologist* 30(11), 1103–1126.

Crawford, S. E. S., and Ostrom, E. (1995). A grammar of institutions. *Am. Political Sci. Rev.* 89(3), 582–600.

Hess, C. (1999). *A Comprehensive Bibliography of Common Pool Resources* (CD-ROM). Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington (www.indiana.edu/~iascp/library.html).

Holland, J. H. (1995). *Hidden Order. How Adaptation Builds Complexity*. Addison-Wesley, Reading, Massachusetts.

Jacob, F. (1977). Evolution and tinkering. *Science* 196(4295) (June 10), 1161–1166.

Landau, M. (1969). Redundancy, rationality, and the problem of duplication and overlap. *Public Admin. Rev.* 29(4), 346–358.

Landau, M. (1973). Federalism, redundancy, and system reliability. *Publius* 3(2), 173–196.

McGinnis, M. (ed.). (1999). *Polycentric Governance and Development: Readings from the Workshop in Political Theory and Policy Analysis*. University of Michigan Press, Ann Arbor.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, New York.

Ostrom, E., Agrawal, A., Blomquist, W., Schlager, E., and Tang, S. Y. (1989). *CPR Coding Manual*. Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington.

Ostrom, E., Gardner, R., and Walker, J. M. (1994). *Rules, Games, and Common-Pool Resources*. University of Michigan Press, Ann Arbor.

Ostrom, V. (1997). *The Meaning of Democracy and the Vulnerability of Democracies: A Response to Tocqueville's Challenge*. University of Michigan Press, Ann Arbor.

Schlager, E. (1994). Fishers' institutional responses to common-pool resource dilemmas. In *Rules, Games, and Common-Pool Resources* (E. Ostrom, R. Gardner, and J. M. Walker, eds.), pp. 247–265. University of Michigan Press, Ann Arbor.

Singleton, S. (1998). *Constructing Cooperation: The Evolution of Institutions of Co-management in Pacific Northwest Salmon Fisheries*. University of Michigan Press, Ann Arbor.

Tang, S. Y. (1992). *Institutions and Collective Action: Self-Governance in Irrigation*. ICS Press, San Francisco.

Wilson, J. A., Low, B., Costanza, R., and Ostrom, E. (1999). Scale misperceptions and the spatial dynamics of a social-ecological system. *Ecological Economics* 31(2), 243–257.

Yoder, R. D. (1991). Peer training as a way to motivate institutional change in farmer-managed irrigation systems. In *Proceedings of the Workshop on Democracy and Governance. Decentralization: Finance & Management Project Report*, pp. 53–67. Associates in Rural Development, Burlington, Vermont.

COMPETITION, INTERSPECIFIC

Bryan Shorrocks
Centre for Biodiversity and Conservation, University of Leeds

GLOSSARY

carrying capacity Maximum number of organisms that can be supported by a given habitat, based on the amount of resources available (such as food, nutrients, shelter, and space).

exploitation competition Interaction among two or more species that use a common resource that is limited, in which one species benefits more than the other.

interference competition Interaction among two or more species that use a common resource that is not limiting, in which one species is harmed by having its access to the resource restricted.

resource partitioning Ecological arrangement in which two or more species use different, nonoverlapping resources in a given habitat, such as warblers foraging for insects in different locations within a tree or canopy.

COMPETITION IS A RECIPROCAL NEGATIVE INTERACTION (affecting the numbers of individuals or some characteristic such as body size) between two species that is brought about by two individuals requiring the same resource, which is frequently in short supply. The intensity of the competition is dependent on the density of the competing populations.

I. INTRODUCTION

Interactions between species can be classified and defined in various ways. A very useful method of classification is to consider the "effect" that individuals of one species have on the population growth of another species and vice versa. We ask the question: In the presence of species A (+) does species B, (1) increase its numbers (+), (2) not change its numbers (0), or decrease its numbers (−) relative to when species A is absent (−)? The same question is asked of species A in the presence of species B. The answers are conveniently summarized in Table I.

Because of the symmetry there are only six types of interaction. These are frequently called (1) 00 neutralism, (2) +0 commensalism, (3) +− predator/prey, parasite/host, or herbivore/plant interactions, (4) 0− amensalism, (5) −− competition, and (6) ++ symbiosis. This article focuses on interactions of types (4) and (5) between individuals of different species. Amensal-

TABLE I

A Classification of Two Species Interactions

		Effect of A on B		
		+	0	−
	+	++	+0	+−
Effect of B on A	0	0+	00	0−
	−	−+	−0	−−

ism is just an extreme example of asymmetric competition. This is competition, usually between a pair of species, in which the adverse effect of one species on the other is much greater than the reciprocal effect.

The negative effects on numbers are produced by individuals directly competing for essential resources (frequently food and space) that are in short supply. If the negative effect is caused by not getting enough of the limited resource, we call it exploitation competition. If the resource is not limiting, but individuals nonetheless harm each other in the process of obtaining it, we call it interference competition. An ecological situation can arise where two species appear to show the reciprocal negative effects associated with interspecific competition, but this is in fact the result of predation by a third species. This is called apparent competition.

Consider the situation shown in Fig. 1. A single species of predator attacks two species of prey. The predator/prey interactions are of −+ type and therefore both species are adversely affected by the predator, and the predator is positively affected by both species of prey. This means that the positive effect that prey 1 has on the predator will, in turn, increase the negative effect on prey 2, and vice versa. The overall consequence of this is that there will appear to be a −− interaction between prey 1 and prey 2, even if they are not competitors for any essential and limiting resource.

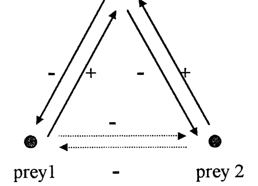


FIGURE 1 Diagram showing apparent competition between two prey species. The solid lines indicate direct interactions between prey and predator, and the dotted lines show indirect (apparent) interactions between the two prey species.

How have ecologists studied competition in order to see the patterns that this ecological process produces? In general, there are four approaches (in common with all studies in ecology and, in fact, all science). These are (1) mathematical models (equations, graphs, or computer simulations), (2) laboratory experiments (laboratory models), (3) field experiments (field models), and (4) field observations. Each approach has its advantages and disadvantages. As you progress from (1) through (2) and (3) to (4), the unit of study becomes more complex but more realistic. Approach (1) helps us to understand a process (or system) and suggests what kind of (2) and (3) studies may be useful. In turn, these may direct us towards particular kinds of field observation that will help confirm or reject the patterns suggested by (1), (2), and (3). In fact the study of competition (its historical development) progressed in part from (1) to (4) and it is convenient to do the same in this account of interspecific competition. The remaining sections illustrate these four approaches with three examples.

II. COMPETITION MODELS

A. The Lotka–Volterra Model

Vito Volterra was an Italian mathematician who, inspired by his son-in-law Umberto D'íAncona (an eminent hydrobiologist), published work on competitive interactions in the 1920s and 1930s. A. J. Lotka was an American mathematician who published on similar topics, also in the 1920s and 1930s. Their names are associated with a basic model of interspecific competition that has had a major impact on how ecologists think about this species interaction.

The Lotka–Volterra model of interspecific competition is based on two other models of population growth, the exponential growth model and the logistic sigmoid growth model. Let us look first at exponential growth, that is, when there is no competition of any kind. Under such conditions the rate at which numbers change with time (dN/dt) can be represented by a per capita (per head) rate of increase multiplied by the number of individuals (N) in the population. The rate of increase used is the intrinsic rate of natural increase (r) and is equal to the difference between the per capita birth rate and the per capita death rate. In mathematical symbols this can be written as

$$\frac{dN}{dt} = rN$$

In this simple model, the rate of increase does not change with density, that is, there is no density dependence and growth is unlimited. However, in most real populations, growth is a function of density (fN) and therefore the growth equation should be written as

$$\frac{dN}{dt} = rN(fN)$$

The simplest relationship that this function could have with density is shown in Fig. 2. We know that the end points of this relationship are quite sensible. At very low population size ($N \cong 0$), fN must $\cong 1$ because this would give unlimited exponential growth. At the equilibrium population size or carrying capacity, $N = K$, $fN = 0$ and growth stops ($dN/dt = 0$). Joining these two known points by a straight line assumes the simplest relationship possible and is not based on any empirical evidence. (Scientists, including ecologists, always develop their descriptions of the natural world by initially assuming the simplest relationships. Only when this proves inadequate do they make the explanation more complex. See Francisco Ayala's experiments discussed in Section III,D.) The slope of the line is equal to $1/K$ and the equation of this straight line is therefore $fN = 1 - N/K$. Substituting the value of fN into the previous equation now produces

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)$$

which is the well-known logistic equation that produces sigmoidal limited population growth. It is important to remember that the term in brackets represents intraspecific density-dependent regulation. The inhibitory effect of one individual on population growth is $1/K$ and the inhibitory effect of N individuals on population growth is N/K.

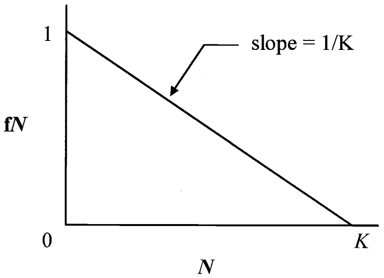


FIGURE 2 The density-dependent function used to produce the logistic equation.

Now let us consider two species, each of which when living separately has its growth described by a logistic equation. However, when living together this logistic equation must be modified. In the following equations the two species are denoted by the subscripts 1 and 2. Let us consider first of all the population growth of species 1 when it competes with species 2. In terms of its own population growth, as described by the logistic equation, the inhibitory effect of one individual of species 1 on its own population growth is $1/K_1$, where K_1 is the carrying capacity of species 1. The inhibitory effect of every individual (N_1) of species 1 on its own population growth is therefore N_1/K_1. If species 2 was identical to species 1 in its effect on the population growth of species 1, then we could simply write the effect of every individual of species 2 on the population growth of species 1 as N_2/K_1. However, this is unlikely. It is more realistic to imagine that each individual of species 2 would have a greater or lesser effect on the population growth of species 1, and therefore we have to multiply N_2/K_1 by a constant that expresses this difference. This constant (usually called the competition coefficient and denoted by α_{12}) describes the relative effect of an individual of species 2 compared to an individual of species 1. If an individual of species 2 was identical (from an ecological point of view), then $\alpha_{12} = 1$. Similarly, the effect of every individual of species 2 on its own population growth will be N_2/K_2 and the effect of every individual of species 1 on the population growth of species 2 will be N_1/K_2 multiplied by α_{21} (note the reversal of the subscripts).

To understand the competition coefficients (α_{12} and α_{21}) more clearly, let us take an imaginary example in which competition is solely for a single food resource. Let us imagine that two species of desert mouse eat seeds. Mouse 1 consumes 100 seeds per day, and mouse 2 consumes 200 seeds per day. In terms of resource units (seeds), an individual of species 1 = half an individual of species 2, or two individuals of species 1 = one individual of species 2. Therefore, $\alpha_{12} = 2$ (the effect of adding one more individual of species 2 is like adding two more individuals of species 1) and $\alpha_{21} = 1/2$.

We can now rewrite the two logistic growth equations for the two species (which incorporate intraspecific competition) to take into account interspecific competition:

$$\frac{dN_1}{dt} = r_1N_1\left[1 - \left(\frac{N_1}{K_1}\right) - \left(\frac{\alpha_{12}\cdot N_2}{K_1}\right)\right]$$

$$= r_1N_1\left[\frac{K_1 - N_1 - \alpha_{12}N_2}{K_1}\right]$$

$$\frac{dN_2}{dt} = r_2N_2\left[1 - \left(\frac{N_2}{K_2}\right) - \left(\frac{\alpha_{21}\cdot N_1}{K_2}\right)\right]$$

$$= r_2N_2\left[\frac{K_2 - N_2 - \alpha_{21}N_1}{K_2}\right]$$

This system of two species (the Lotka–Volterra model of two-species competition) will be at equilibrium when $dN/dt = dN/dt = 0$, that is, when both species are at their carrying capacities and population growth has stopped. Just considering for the moment species 1, this means that

$$r_1N_1\left[\frac{K_1 - N_1 - \alpha_{12}N_2}{K_1}\right] = 0$$

Since r_1, N_1, and K_1 cannot $= 0$ (mathematically possible but ecologically uninteresting if species 1 cannot reproduce, is not present, or cannot exist in that habitat), this means that for species 1, $K_1 - N_1 - \alpha_{12}N_2 = 0$ at equilibrium. Rearranging this equation we get $N_1 = K_1 - \alpha_{12}N_2$, which is the equation of a straight

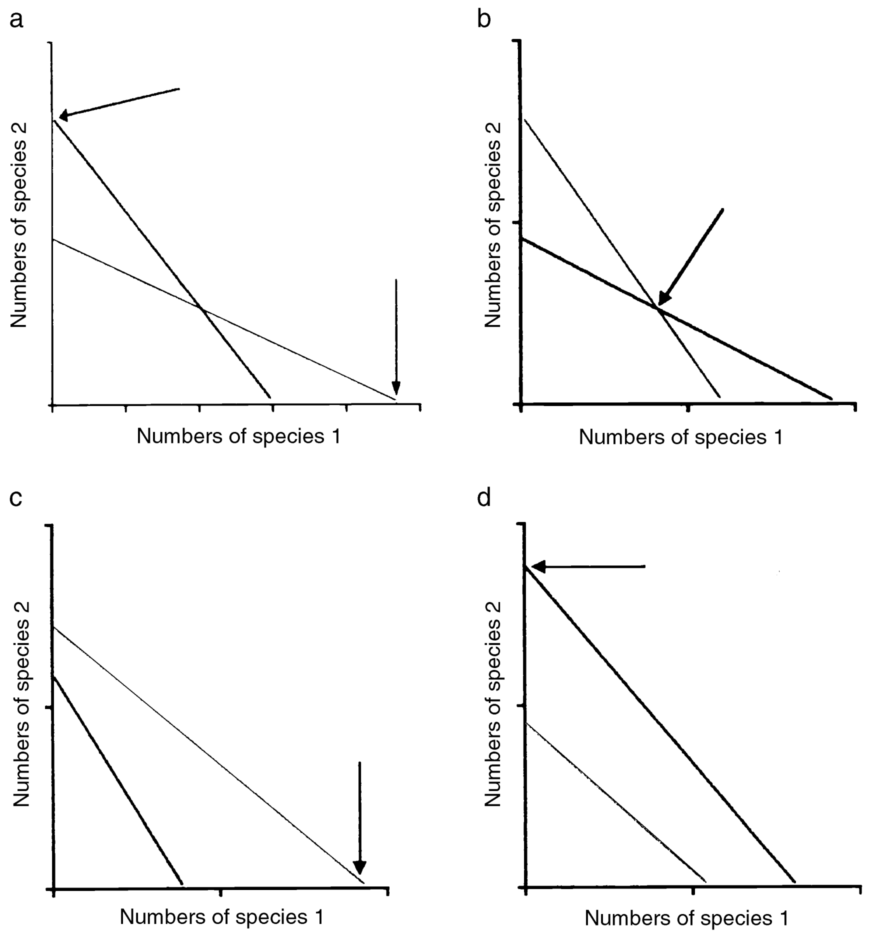


FIGURE 3 Graphical representation of the four predicted outcomes of the Lotka–Volterra model. Arrows indicate equilibrium points. The thin line = species 1 zero isocline and the thick line = species 2 zero isocline.

line and is called the species 1 zero isocline. On a graph of the numbers of species 1 (N_1) plotted against the numbers of species 2 (N_2), it represents all those combinations of N_1 and N_2 for which the population growth of species 1 (dN_1/dt) = 0.

Following a similar procedure for species 2 will produce the species 2 zero isocline, which is $N_2 = K_2 - \alpha_{21}N_1$. On a graph of the numbers of species 2 (N_2) plotted against the numbers of species 1 (N_1), it represents all those combinations of N_1 and N_2 for which the population growth of species 2 (dN_2/dt) = 0.

There are four ways that these two straight lines can be placed on a graph of N_2 versus N_1, and these are illustrated in Fig. 3. These are known as phase plane graphs/diagrams. Each point on the graph represents a joint density of the two species. To each point in this graph we can attach an arrow, or vector, whose length and direction will indicate the dynamics of the system. Below its zero isocline, a species can increase in numbers. Above its zero isocline, a species will decline in numbers. Exactly on its zero isocline, a species will neither increase nor decrease. This means that the four graphs predict different outcomes for interspecific competition. In (a) either species can win depending on which area between the zero isoclines the population trajectory reaches. This depends on starting numbers and/or relative growth rates (r_1 and r_2). The point were the two lines cross represents an unstable equilibrium point. In (b) the two species coexist at the values of N_1 and N_2 were the two lines cross. This represents a stable equilibrium point. In (c) species 1 always wins and in (d) species 2 always wins.

Laboratory examples of one species winning (c and d) and coexistence (b) are provided by G. F. Gause's experiments (see Section III,A). Examples of one species winning (c and d) and either species winning (a) are provided by *T. Park's* experiments (see Section III,B). It is important to make two comments about the Lotka–Volterra model. First, it contains hidden variables. The competition coefficients (α) are simply conversion factors to allow one species to be represented in the other species numerical equivalents. Species may compete for food and space and poison each other, but all of these mechanisms of competition are summarized and hidden within α. Second, because the real world (unlike most laboratory environments) is heterogeneous and patchy, even (a), (c), and (d) may result in coexistence. If within a habitat patch (a), (c), and (d) apply, there may only be one species per patch. But looking over many patches, both species will be seen to apparently "coexist." This is the difference between what ecologists call "local" and "regional" coexistence.

What do these graphs in Fig. 3 imply for competition and coexistence? Let us take graph (a) as an example (species 1 or species 2 wins depending on initial concentrations and/or the rates of increase). Using the end points of the lines, this graph can be defined as

$$\alpha_{12} > K_1/K_2 \qquad \alpha_{21} > K_2/K_1$$

or by dividing the two sides of the left-hand inequation by K_1 and the two sides of the right-hand inequation by K_2,

$$\alpha_{12}/K_1 > 1/K_2 \qquad \alpha_{21}/K_2 > 1/K_1$$

In competition terms this means that the

$$\begin{array}{l}\text{effect of one}\\ \text{individual of}\\ \text{species 2 on}\\ \text{species 1}\end{array} > \begin{array}{l}\text{species 2}\\ \text{on}\\ \text{species 2}\end{array} \qquad \begin{array}{l}\text{species 1}\\ \text{on}\\ \text{species 2}\end{array} > \begin{array}{l}\text{species 1}\\ \text{on}\\ \text{species 1}\end{array}$$

In other words, each species reduces the *dN/dt* of the other species more than its own *dN/dt*, which we can summarize as

$$2 \text{ on } 1 > 2 \text{ on } 2 \qquad 1 \text{ on } 2 > 1 \text{ on } 1$$

Interspecific effects for both species are stronger than intraspecific effects.

For graph (b) (coexistence) these inequalities are

$$2 \text{ on } 1 < 2 \text{ on } 2 \qquad 1 \text{ on } 2 < 1 \text{ on } 1$$

which is the opposite of graph (a). Here each species reduces its own *dN/dt* more than that of the other species, and therefore interspecific effects for both species are weaker than intraspecific effects. This case has always intrigued ecologists, and the traditional explanation as to why interspecific effects for both species should be weaker than intraspecific effects is "resource partitioning" (see Section II,B). However, there are other mechanisms that will produce this effect (see Section II,C).

For graph (c) (where species 1 always wins) the inequalities are

$$2 \text{ on } 1 < 2 \text{ on } 2 \qquad 1 \text{ on } 2 > 1 \text{ on } 1$$

which implies that the two species now do different things. Species 2 reduces its own *dN/dt* more than that of species 1, and species 1 reduces the *dN/dt* of species 2 more than its own *dN/dt*.

For graph (d) (where species 2 always wins) the inequalities are

$$2 \text{ on } 1 > 2 \text{ on } 2 \qquad 1 \text{ on } 2 < 1 \text{ on } 1$$

and we have the opposite of graph (c).

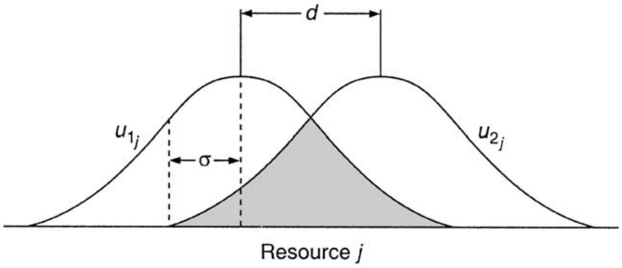


FIGURE 4 Resource utilization curves used in the MacArthur model. *D* is the difference between the means of the two normal distributions and σ is the standard deviation for the two normal distributions.

B. Resource Utilization Curves

The Lotka–Volterra competition model predicts that two potential competitors will coexist if they each affect the growth of the other more than they affect their own growth. That is, intraspecific effects are greater than interspecific effects for both species. This is summarized by the inequations

$$\alpha_{12}/K_1 < 1/K_2 \qquad \alpha_{21}/K_2 < 1/K_1$$

Traditionally this led to the concept of resource separation (different niches) as a mechanism of coexistence. Clearly, if two species use different resources they will interact (compete) only with their own species, and therefore the preceding inequalities will be true. But partial resource partitioning has also been used widely as an explanation for coexistence. Field observations in particular (see Section V) have frequently found partial resource separation between species and accumulated these differences over several “niche dimensions” to convince the reader that enough separation exists to “explain” the coexistence observed. Unfortunately, most of these field studies do not calculate α_{12}, α_{21}, K_1, or K_2 and therefore we do not know if sufficient separation exists between the species to make $\alpha_{12}/K_1 < 1/K_2$ and $\alpha_{21}/K_2 < 1/K_1$. However, one attempt to quantify the “amount” of separation required is provided by a simple model, initially developed by Robert MacArthur (1972).

Imagine two species competing for a resource that can be visualized as varying continuously in one dimension, for example, insects or seeds of different sizes or foraging sites up a tree. Imagine also that the two species’ utilization of this resource follows a unimodal distribution, with a preferred area of use in the center and less preferred resource items on either side. Figure 4 shows this idea for two species using a resource *j*, in which their resource utilization curves (u_{1j} and u_{2j}) are described by a normal distribution. Each species uses a different part of the resource, with some overlap. The distance between the two means is *d*, and σ is the standard deviation. The two species are assumed to have the same carrying capacity and the intensity of competition is related to the area of overlap between the two distributions. With two normal distributions this is given by

$$\alpha = e^{-d^2/4\sigma^2}$$

and the relationship between α and d/σ is shown in Fig. 5. For stable coexistence between two species,

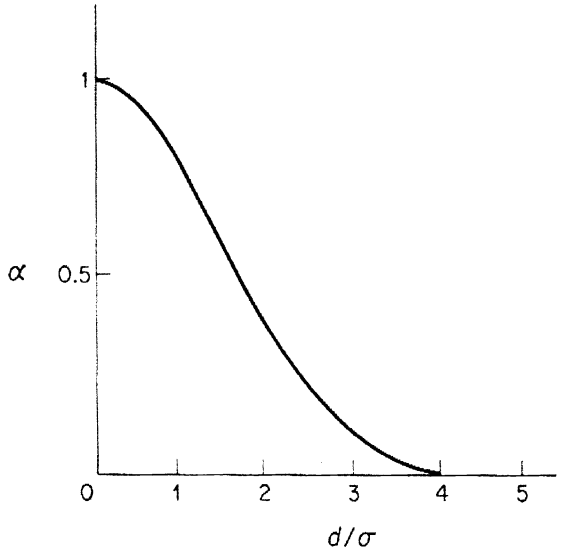


FIGURE 5 Relationship between α and the ratio d/σ predicted from the MacArthur model.

d/σ should be approximately greater than unity, that is, d (the distance between means) should be greater than σ (the standard deviation of resource utilization). A field example that agrees with this amount of limiting similarity between species is provided by the foraging height relationships within antbirds (*Myrmotherula*), shown in Fig. 6. Unfortunately, it has been pointed out that $d/\sigma > 1$, and promoting coexistence may be a property of this type of model only with a one-dimensional resource. With competition in several dimensions and alternative resource utilization curves, lower values of d/σ may be compatible with stable coexistence.

C. The Aggregation Model

Many insects exploit resources that are patchy, consisting of small, separate units, and that are ephemeral in the sense that they persist for only one or two generations. Such resources can include fruit, fungi, sap flows, decaying leaves, flowers, dung, carrion, seeds, dead wood, and small bodies of water held in parts of terrestrial plants (phytotelmata). This general view of insect ecology inspired the aggregation model of competition (Shorrocks *et al.*, 1984; Atkinson and Shorrocks, 1981), which allows a competitively inferior species to survive in probability refuges. These are patches of resource (a single fungus, fruit, etc.) with no or a few superior competitors, that arise because the competing stages (usually larvae) have an aggregated distribution across the patches. These probability refuges are a permanent feature of such systems because patches, such as fungi, are ephemeral and aggregation increases mean crowding. Regional population density is limited by strong intraspecific competition in patches with high local density while low-density patches still exist (e.g., population size within a wood is limited by high density in some fungi, while other fungi still contain no or a few individuals). As with resource partitioning, coexistence is promoted because aggregation of the superior species increases its intraspecific competition and reduces interspecific competition.

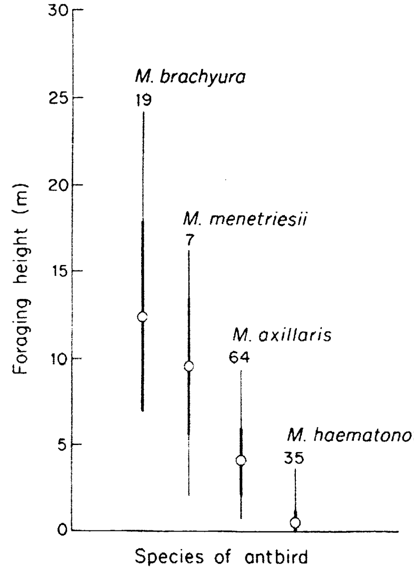


FIGURE 6 Foraging height of antbirds (Formicariidae). Notice that the means are separated by about one standard deviation, as predicted by the MacArthur model.

In the aggregation model, the eggs of both insect species are independently distributed over the patches according to a negative binomial distribution, which has an exponent, k, inversely related to the degree of intraspecific aggregation. The use of the negative binomial and the assumption of independence have been justified for drosophilid flies. In the first version of the model, the parameter k (level of aggregation) was constant and independent of density. This is not valid for real populations, but relaxing this assumption does not prevent coexistence. Within each patch, competition is modeled by a difference form of the Lotka–Volterra equations.

The predictions of the aggregation model are that with k of the negative binomial <1 (strong aggregation), it is virtually impossible for the competitively "superior" species to eliminate the competitively "inferior" species. Figure 7 shows the model's results as a graph of "critical α" against k of the negative binomial. Critical α is the competition coefficient that the "superior species" must have in order to exclude the "inferior species." Also shown on the graph are distributions of α and k for drosophilid flies. For these flies it is clear that k of the negative binomial is usually <1 and that competition coefficients are not sufficiently large to prevent competitive exclusion. For many animals exploiting ephemeral and patchy resources, this model therefore provides a viable alternative to traditional resource partitioning as an explanation for the coexistence of species. The two-species model has been extended to a many-species model (Shorrocks and Rosewell, 1987) and predicts average group sizes of about seven species coexisting on identical resources.

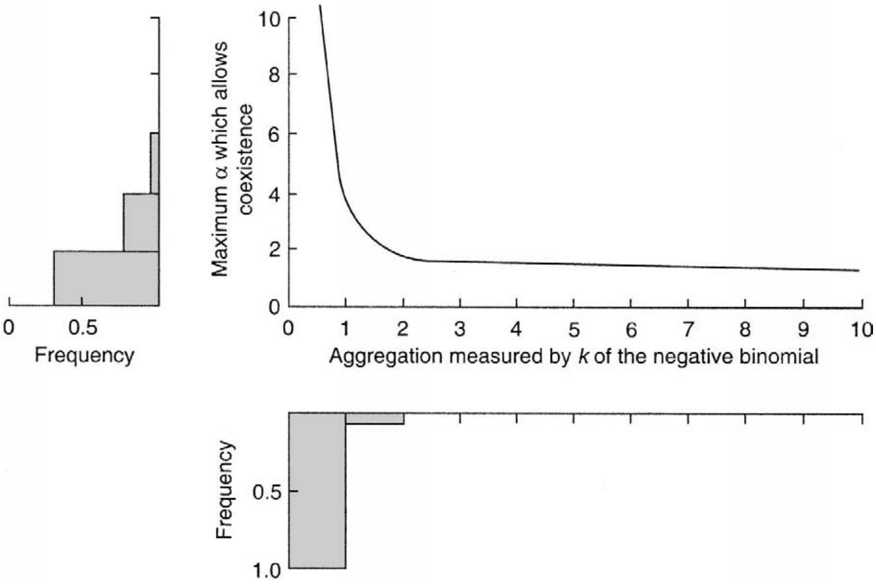


FIGURE 7 Relationship between "critical α" and aggregation (*k* of the negative binomial) predicted by the aggregation model.

III. LABORATORY EXPERIMENTS

A. Yeast and *Paramecium*

In the early part of this century, G. F. Gause carried out a series of simple experiments with yeasts and protozoa that are instructive, are of historic interest, and set the experimental protocol for many of the so-called "bottle" experiments conducted in the 1940s, 1950s, and 1960s. His early experiments are detailed in his classic book *The Struggle for Existence* (Gause, 1934). He was mainly interested in testing the Lotka–Volterra model that had been put forward as a description of the competitive process, between species, a few years earlier.

The experimental design was simple. He grew two species, each alone and then together in mixed culture, to determine the course of competition under laboratory conditions. In his first experiments he used two species of yeast, which he called *Saccharomyces* and *Schizosaccharomyces*. However, we are not now sure of the exact taxonomic identity of the two species. He grew these species on an extract of brewer's yeast, water, and sugar. The growth medium was not renewed, but growth ceased before the sugar resource was exhausted, apparently because of the accumulation of alcohol. These experiments with yeast are therefore an example of interference competition rather than exploitation competition. The main interest of these experiments lies in the calculation of the competition coefficients (α), which Gause did by estimating all the variables in the Lotka–Volterra competition equations except α. The intrinsic rate of natural increase (r_1 and r_2) and the carrying capacity (K_1 and K_2) were estimated by fitting a logistic equation to the single-species growth curves. From the mixed cultures, he estimated both the numbers of each species (N_1 and N_2) and the growth rates (dN_1/dt and dN_2/dt). By rearranging the Lotka–Volterra equations he obtained

$$\alpha_{12} = \frac{K_1 - \dfrac{dN_1/dt \cdot K_1}{r_1 \cdot N_1} - N_1}{N_2}$$

$$\alpha_{21} = \frac{K_2 - \dfrac{dN_2/dt \cdot K_2}{r_2 \cdot N_2} - N_2}{N_1}$$

Since all quantities on the right-hand side are known, he could estimate the competition coefficients (α_{12} and α_{21}). Since N_1, N_2, dN_1/dt, and dN_2/dt all vary with time, Gause selected three occasions during his experiments to estimate α. He then took the mean of these three calculations. His experiments were also repeated under two environmental conditions, which he called anaerobic (yeasts grown in test tubes) and aerobic (yeasts grown in flasks). Since α is a measure of the effect that

TABLE II

Competition Coefficients (α) Calculated from Yeast Population Experiments and Relative Alcohol Production, under Both Anaerobic and Aerobic Conditions

	Experiments	Alcohol
Anaerobic		
α_{12}	3.15	2.19
α_{21}	0.44	0.46
Aerobic		
α_{12}	1.25	1.25
α_{21}	0.85	0.80

one individual of species 1 has on the population growth of species 2 (and vice versa), and with these species competition is thought to be mainly the result of alcohol poisoning, it should be possible to calculate α directly from the relative production of alcohol by the two species. Table II shows the values of α calculated from population experiments and relative alcohol production, under both anaerobic and aerobic conditions.

Two important points come out of these yeast experiments. First, the competition coefficients are not constant; they vary (quite markedly) with different environmental conditions. Second, relative alcohol production and its effect on growth are sufficient to explain the interference competition between these two species (particularly under aerobic conditions). Therefore, in this experiment, the competition coefficients are relatively simple, pure quantities and $\alpha_{12} \cong 1/\alpha_{21}$.

In his later and perhaps more famous experiments, Gause used two species of single-celled protozoa called *Paramecium*. Initially he used what he called *P. caudatum* (1) and *P. aurelia* (2), although once again we are not sure about the precise taxonomy because these "species" are now known to be closely related groups of species. The *Paramecium* were grown in 5 cm^3 of Osterhaut's medium (a mixture of salts) in test tubes at 26°C. Each day fresh food was added to the test tubes. The food was a bacterium grown on agar plates and added to the cultures using a platinum loop (to standardize the amount given). The bacterium used did not grow in Osterhaut's medium. Each experiment was started with 20 *Paramecium* individuals on Day 0, and Gause estimated the number of individuals each day by sampling 10% of the medium. Since an individual of *P. aurelia* occupies only 39% of the volume of a single *P. caudatum*, he estimated the volume of a typical cell for each species and converted the estimated numbers to biomass (in volume). Following the same protocol as the yeast experiments, the two species were grown alone (four cultures of *P. caudatum* and three cultures of *P. aurelia*) and in mixed culture (three cultures); Fig. 8 shows the mean volumes, over time, for these experiments. In single culture, both species appear to grow to an equilibrium at around $K = 200$. In mixed culture, *P. caudatum* appears to suffer more from interspecific competition than does *P. aurelia*, although the latter also grows less well. Using the same procedure as for the yeast experiments, Gause was able to estimate the competition coefficients (α_{12} and α_{21}), and these are given in Table III.

Notice that in the first few days the competition coefficient for the effect of *P. aurelia* on *P. caudatum* is negative. This implies that the effect of *P. aurelia* is actually positive at this stage. Given that $K_1 = K_2 = 200$, $\alpha_{12} = 1.64$, and $\alpha_{21} = 0.61$, then we can calculate $K_1/\alpha_{12} = 122$ and $K_2/\alpha_{21} = 328$ and therefore construct the Lotka–Volterra phase-plane for this experiment. This is shown in Fig. 9, along with the two-species population trajectory for the results in Fig. 8. These experiments with *Paramecium* appear to broadly confirm the predictions of the Lotka–Volterra model of interspecific competition, although the mortality imposed by the sampling procedure (10%) and the premature termination of the experiments result in the predicted equilibrium not being reached in 25 days. However, later experiments in which Osterhaut's medium was buffered to a pH of 8 and the daily food concentration was reduced resulted in *P. caudatum* being excluded from the cultures in just 16 days.

In another set of experiments using *P. aurelia* and *P. bursaria*, Gause obtained population trajectories that suggested coexistence, although again the experiments were not run for long enough to reach an actual equilibrium. However, irrespective of the starting numbers, the two-species population trajectories converge on the same joint densities. *Paramecium bursaria* is interesting since it contains within its cell body unicellular green plants. The suggestion has been made that *P. bursaria* was able to exploit the bacterial food that had settled to the bottom of the experimental tube. This area of

TABLE III

Competition Coefficients Calculated from Gause's *Paramecium* Experiments

	α_{12}	α_{21}
First days	−1.00	+0.50
Greater than fifth day	+1.64	+0.61

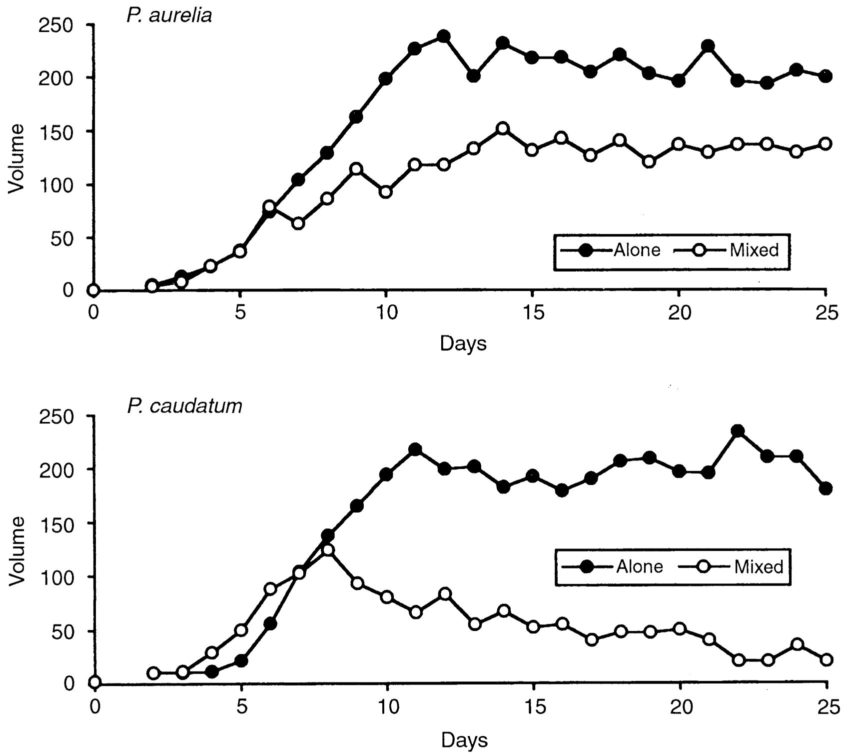


FIGURE 8 Growth in "volume" of *Paramecium caudatum* and *P. aurelia* in both single and mixed culture (from Gause, 1934). Volumes are converted from number of individuals per 0.5 cc in the culture. Since the volume of *P. caudatum* was set at 1.00 and that of *P. aurelia* at 0.39, the volume for *P. caudatum* is equal to the number of individuals per 0.5 cc.

the microhabitat is depleted of oxygen (by the bacteria) and cannot be used *by P. aurelia*. However, *P. bursaria* carries its own oxygen supply (produced by its unicellular plants) and can use this part of the bacterial resource. Heterogeneity within this simple habitat, produced by the bacterial food, may have been sufficient to allow resource partitioning.

B. Flour Beetles

In a series of studies, Thomas Park carried out competition experiments between two species of flour beetle, *Tribolium confusum* (1) and *T. castaneum* (2). Both are frequently found as stored-product pests and they can be conveniently raised under laboratory conditions. Populations were kept in glass tubes with 8 g of medium (95% flour and 5% yeast). All adults, pupae, and larvae were counted every 30 days and both single- and mixed-species populations were kept. An interesting aspect of the experimental design, which proved instructive, was that populations were maintained at a number of temperatures (24°, 29°, and 34°C) and relative humidities (30% and 70%), and each type of population (single or mixed), under all six environmental conditions, was replicated 30 times.

1. Single-Species Results

Apart from *T. castaneum* at 24°C and 30% humidity, all single-species populations persisted, and Table IV shows the average population densities during the course of the experiment. The value of 3, for *T. castaneum* at 24°C and 30% humidity, is the average population density until extinction.

Using these average densities as an indication of success, Park compared the relative performance of these two species (alone) for each temperature/humidity combination. The results of this comparison are shown in Table V, where an asterisk indicates a statistically significant difference between the two species. Except for 34°C and 70% humidity, each of the physical environments significantly favors one species over another. These results are due to intraspecific competition and the effect of the physical environment on it.

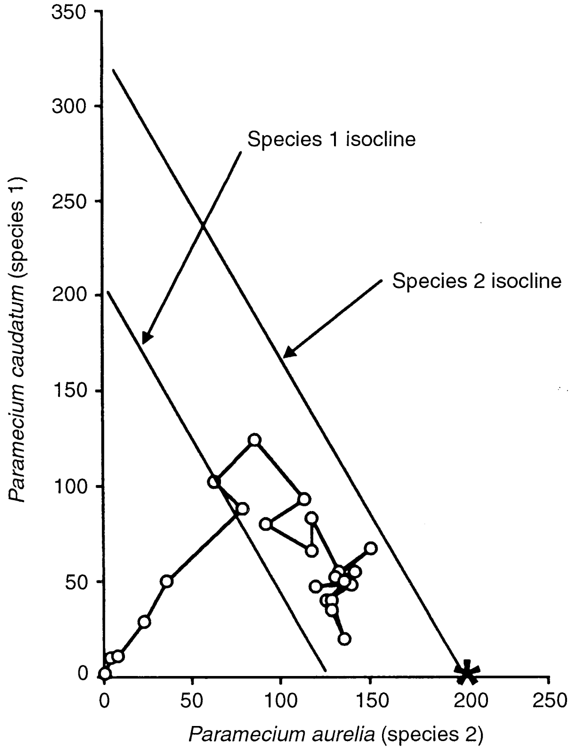


FIGURE 9 Phase-plane graph showing the experimental trajectory from Fig. 8. Numbers are actually the "volume" of each species. Intercepts (and therefore zero isoclines) are worked out from Gause's experimental data (see text). The predicted equilibrium is indicated by an asterisk.

2. Two-Species Results

Park found that while single-species populations persisted (with the exception of *T. castaneum* at 24°C and 30% humidity), mixed-species populations did not. In each population only one species persisted and the other was eliminated. However, unlike most previous "bottle" experiments (e.g., Gause's experiments), it was not always the same species that was eliminated, even under the same environmental conditions. This "indeterminate" outcome of competition is one of the predicted possible outcomes of the Lotka–Volterra model. Table VI shows the percentage of replicates that *T. castaneum* won.

These experiments illustrate two interesting and possibly universal features of competition. First, the outcome of two-species competition can sometimes be predicted from the performance of each species on its own (i.e., 29°C/70% and 24°C/30%), but also *sometimes it cannot* (i.e., 34°C/70% and 24°C/70%). Second, the outcome of competition between two species can vary along an environmental gradient. Park's experiments go from hot-humid conditions (34°C/70%) to cool-dry conditions (24°C/30%). At the hot-humid end, *T. castaneum* eliminates *T. confusum* each time. At the cool-dry end, *T. confusum* eliminates *T. castaneum* each time. In the middle region either species can win, but *T. castaneum* wins more often toward the hot-humid end, and *T. confusum* wins more often toward the cool-dry end. If we imagine that the isoclines in the Lotka–Volterra phase-plane graphs can move their position depending on environmental conditions, then these *Tribolium results are still understandable in terms of the Lotka–Volterra model.* Figure 10 presents a series of phase-plane graphs, going from hot-humid (left) to cool-dry (right). The arrow(s) in each diagram indicate the equilibrium point. Notice that the species predicted to win

TABLE IV

Average Densities of Beetles Observed in Park's Single Species Experiments

	34°C	29°C	24°C
T. confusum (1)			
70%	41	33	29
30%	24	30	30
T. confusum (2)			
70%	30	50	45
30%	10	19	3

TABLE V

Comparison of Park's Single-Species Beetle Experiments

	34°C	29°C	24°C
70%	1 > 2	2 > 1*	2 > 1*
30%	1 > 2*	1 > 2*	1 > 2*

[a] Asterisks indicate that *T. confusum* and *T. castaneum* maintained significantly different population sizes.

TABLE VI

The Percentage of Replicates That *T. castaneum* Won in Park's Beetle Experiments

	34°C	29°C	24°C
70%	100	86	29
30%	10	13	0

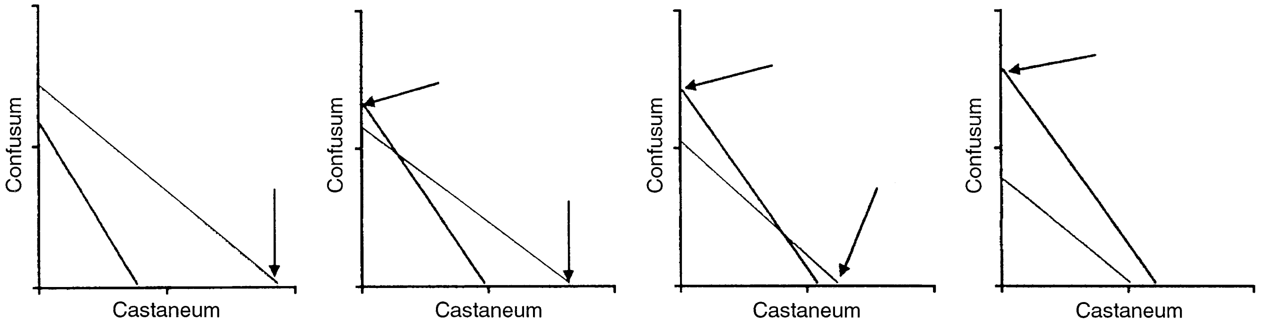


FIGURE 10 A series of phase-plane graphs along an environmental gradient from hot-humid (left) to cool-dry (right). The arrow(s) in each diagram indicate the equilibrium point.

moves from always *T. castaneum* (left), to either species, to always *T. confusum*.

C. Drosophila

In the late 1960s and early 1970s, Francisco Ayala conducted a series of laboratory experiments, each using a pair of *Drosophila* species. Although five species were used in different combinations, the basic design and outcome of all of these experiments were very similar. One experiment, between *D. serrata* and *D. pseudoobscura*, will be sufficient to illustrate the apparent dilemma that emerged.

Drosophila serrata comes from Australia and *D. pseudoobscura* comes from North America. Notice therefore that since neither of these flies comes from the same continent, let alone the same community, Ayala was simply using the flies like analog computers to find out what might happen between flies in real communities.

At 25°C, *D. serrata* eliminates *D. pseudoobscura* in a few generations; at 19°C, *D. pseudoobscura* eliminates *D. serrata*. Ayala performed his experiments at 23.5°C, at which temperature the two species coexisted. Populations were started with 300 adult flies of each species and were maintained in a series of milk bottles with standard *Drosophila* medium (a kind of yeasty porridge). Adult flies feed off the yeasty surface of the medium, female flies lay their eggs onto the surface, and larvae feed within it. Third instar larvae pupate on the surface of the medium or on the side of the bottle. When an apparent equilibrium was reached in the mixed populations, single-species populations were established and maintained under the same conditions. From the population data, and assuming that the Lotka–Volterra model of competition describes what is happening between the two species, Ayala was able to calculate the competition coefficients, α_{12} and α_{21}. He did this in the same manner as Gause, by rearranging the Lotka–Volterra equations:

$$\alpha_{12} = \frac{K_1 - \dfrac{dN_1/dt \cdot K_1}{r_1 \cdot N_1} - N_1}{N_2}$$

$$\alpha_{21} = \frac{K_2 - \dfrac{dN_2/dt \cdot K_2}{r_2 \cdot N_2} - N_2}{N_1}$$

However, by assuming that the two-species populations had reached an equilibrium ($dN_1/dt = dN_2/dt = 0$), these equations can be reduced to

$$\alpha_{12} = \frac{K_1 - N_1}{N_2}$$

$$\alpha_{21} = \frac{K_2 - N_2}{N_1}$$

where K_1 and K_2 equal the equilibrium population densities of each species on their own and N_1 and N_2 equal the equilibrium population densities for each species in the mixed populations.

All the pairs of species that Ayala studied coexisted for many generations, however, in every case he found that

$$\alpha_{12} > \frac{K_1}{K_2} \text{ and } \alpha_{21} > \frac{K_2}{K_1}.$$

This is not the "case for coexistence" in the Lotka–Volterra model of competition. Intriguingly, Ayala did not conclude that the model was incorrect, but rather that the species could coexist even if they utilized the same resources.

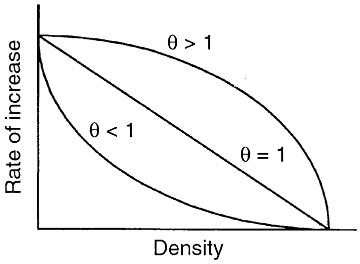


FIGURE 11 The nature of density dependence in Ayala's model.

Of course, the Lotka–Volterra model incorporates the logistic equation, which assumes (as a first approximation) that the relationship between rate of increase and density is a straight line (see Fig. 2). However, Gilpin and Justice examined the relationship between birth/death and density in Ayala's experiments and found that for these *Drosophila* it was frequently a curve. It is not surprising therefore that an apparent conflict had occurred between theory and practice. This led Gilpin and Ayala (1973) to develop a new Lotka–Volterra model that incorporated such "non-linear" relationships. Their new model of competition has four parameters, one more than the Lotka–Volterra model, and explains 95% of the variance in Ayala's experiments. For two species, the model can be written as

$$\frac{dN_1}{dt} = r_1N_1\left[1 - \left(\frac{N_1}{K_1}\right)^{\theta_1} - \left(\frac{\alpha_{12}\cdot N_2}{K_1}\right)\right]$$

$$\frac{dN_2}{dt} = r_2N_2\left[1 - \left(\frac{N_2}{K_2}\right)^{\theta_2} - \left(\frac{\alpha_{21}\cdot N_1}{K_2}\right)\right]$$

The new parameter (θ) describes the "non-linear" nature of density dependence. When $\theta = 1$, the new model is identical with the Lotka–Volterra model. Figure 11 shows the nature of density dependence when $\theta \neq 1$, < 1, and > 1. These modified Lotka–Volterra equations give zero isoclines, on an N_1-by-N_2 graph, that are curves rather than straight lines. This leads to the intriguing possibility that more than one equilibrium may exist for a pair of competing species (Fig. 12).

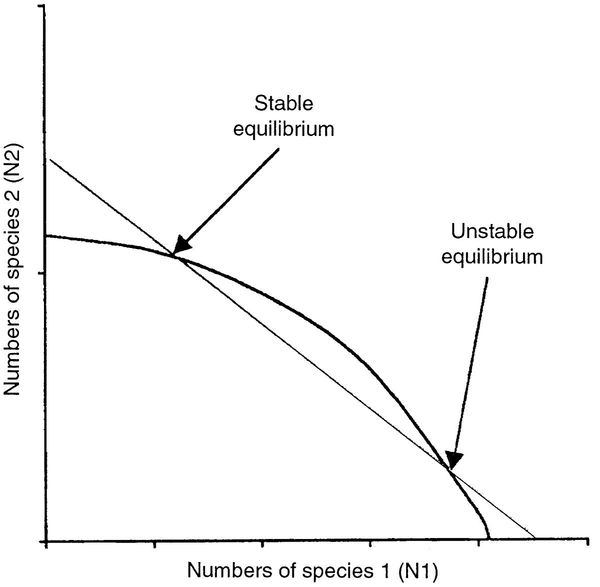


FIGURE 12 Two equilibrium points resulting from curved zero growth lines derived from the Gilpin–Ayala model.

IV. FIELD EXPERIMENTS

A. Barnacles

Barnacles compete for space in the inter-tidal zone of rocky shores. Because they are attached to the rock, they are ideal animals for experimental manipulation in the field. In the 1950s, Joseph Connell conducted a classic set of removal experiments to test for competitive interactions between two species of barnacle, *Chthamalus stellatus* and *Balanus balanoides*. The experiments were carried out on the Isle of Cumbrae in the Firth of Clyde, Scotland.

These two species occupy two separate horizontal zones (with a small area of overlap), with *Chthamalus* (which is more resistant to desiccation) higher up the shore. *Chthamalus* larvae settle on the shore in September/October, whereas *Balanus* settle in April/May. Connell selected eight areas for study, on different parts of the shore, and used old glass lantern slides (10.7 cm × 8.2 cm) as quadrats on which he could mark the position of all individuals of both species of barnacle.

After the settlement of *Balanus* had stopped in early June, a census of the *Chthamalus* was made. Each quadrat was divided into two halves and from one randomly chosen half, all the *Balanus* individuals were removed. Survival of the *Chthamalus* individuals in both halves of the quadrats was then monitored for 12 months. Figure 13 shows three representative results from three of the quadrats. What is quite clear is that survival of *Chthamalus* is much better in those halves of the quadrats without *Balanus*. Direct observation showed that this was due to interference competition from *Balanus* individuals who smothered, undercut, or crushed

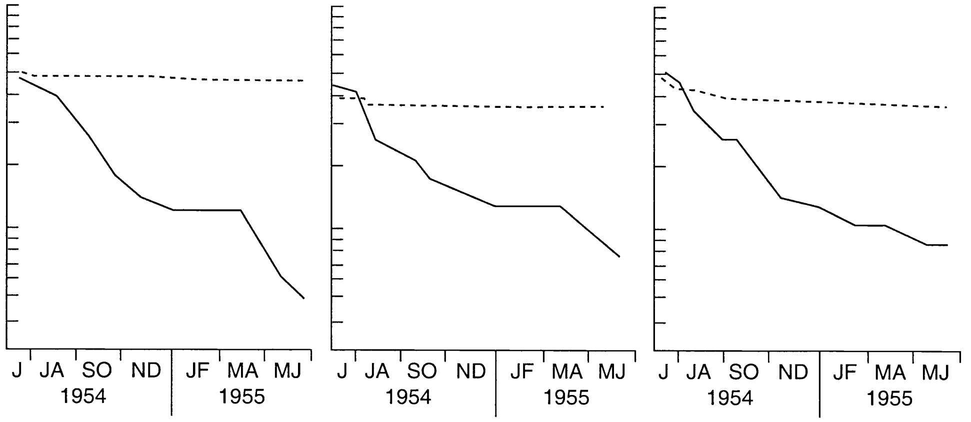


FIGURE 13 Survivorship curves for *Chthamalus stellatus*, with (solid line) and without (dashed line) *Balanus balanoides*.

Chthamalus individuals. Those *Chthamalus* individuals in the *Balanus* half of the quadrat that survived the competition were smaller than those that had not been subjected to this inter-specific competition, illustrating that competition affects fecundity as well as survival.

B. Ants and Mice

Seeds play a major role in the ecology of desert regions and the seeds of annual plants are the primary food of several distantly related taxa of specialized granivores, such as rodents, birds, ants, and beetles. In particular, rodents and ants are very similar in their utilization of seed resources.

In the desert near Tucson, Arizona, experiments demonstrated that ants and rodents take most of the seeds, harvest the same size and species, and collect them from the same microhabitat. Brown and Davidson performed a set of exclusion experiments to demonstrate competition between these two taxa. They established replicated, circular plots (36 m in diameter) in relatively level, homogeneous desert scrub, approximately 60 km northwest of Tucson. Two plots were subjected to each of the following treatments. (1) Plots were fenced to exclude seed-eating rodents. Those rodents present were removed by trapping. (2) Granivorous ants were removed by repeated application of insecticide to individual colonies. (3) Both of the preceding treatments were done. (4) None of the treatments were done (control experiment).

Over a period of several months, the ants were censused five times (colonies counted) and the rodents were censused twelve times (trapping). Table VII shows the "numbers" of ants and rodents in the different treatments at the end of the experimental period. In the "control" experimental plots both ants and rodents showed a decrease in "numbers" compared to their single-taxon plots. Both groups showed a negative effect, which was interpreted as competition for seeds. This suggestion is strongly supported by the fact that on plots from which both ants and rodents were removed, there was a 5.5 times increase in density of seeds. However, there was no difference in seed density between plots with ants + rodents, ants only, and rodents only.

C. *Anolis* Lizards

Anolis lizards are an important component of island terrestrial communities in the eastern Caribbean. In these communities, anoles substitute for the ground-feeding insectivorous bird guild and are major components of the animal biomass. Joan Roughgarden and her colleagues have demonstrated the presence of com-

TABLE VII

Numbers of Ants and Mice in the Exclusion Experiments of Brown and Davidson

	Rodents removed	Ants removed	Control	% increase
Ant colonies	543	—	318	70.8
Rodent numbers	—	144	122	18.0

TABLE VIII

Lizard Introductions to a Caribbean Cay

Grove		Date	*A. gingivinus*	*A. wattsi*
SK	Original	August 1979	196	0
(559) m^2	Start	August 1979	196	103
	End	May 1980	205	14
MK	Original	March 1980	116	0
(263 m^2)	Start	March 1980	75	55
	End	October 1980	116	15
SK	Original	March 1981	194	10
(559 m^2)	Start	March 1981	92	110
	End	October 1981	153	37
LU	Original	March 1981	104	0
(296 m^2)	Start	March 1981	104	48
	End	October 1981	127	7

petition between two species, *Anolis gingivinus* and *A. wattsi*. Both species are insectivorous, territorial as adults, and somewhat arboreal in habit.

On the island of St. Maarten, distributional evidence suggests that there is present-day competition. *Anolis wattsi* occurs only in the central hills of St. Maarten, while *A. gingivinus* occurs throughout the island including the central hills. This observation is interesting because all the relatives of *A. wattsi* on nearby islands occur throughout all elevations and habitats, including sea-level habitats. Two other observations are suggestive of a competitive interaction between these two species. *Anolis gingivinus* shows a lower abundance and shifts its perch position where it co-occurs with *A. wattsi*, relative to when it occurs alone. However, these observational data are not conclusive evidence of competition.

Roughgarden and colleagues therefore carried out some manipulative field experiments on a small cay close to St. Maarten. This very small island was essentially a limestone platform (100 m × 400 m), 15 m above sea level, with several vegetation groves consisting of sea grape, Manchaneel, and perennial grasses. This cay had a resident population of *A. gingivinus*, but lacked any *A. wattsi*. Table VIII shows the results of four experiments carried out over a three-year period in three of the vegetation groves. In the experiments in grove SK (1979) and LU, marked individuals of *A. wattsi* were released and the resident *A. gingivinus* were left undisturbed. Survival of *A. wattsi* was low; most of the introduced individuals disappeared and those that were left did not establish territories in the center of the grove, only at the periphery. In the experiments in grove SK (1981) and MK, between 40% and 50% of the resident *A. gingivinus* were removed prior to introducing *A. wattsi*. In these experiments, survival was twice that in the other experiments and some surviving *A. wattsi* did establish territories in the center of the grove. It seems likely therefore that the restricted distribution of *A. wattsi* (to the central hills) observed on St. Maarten is due to competitive exclusion of *A. wattsi* by *A. gingivinus*.

V. FIELD OBSERVATIONS

A. New England Warblers

Many field studies of competition actually look for resource partitioning (niche separation) that would reduce interspecific competition, relative to intraspecific competition, and therefore promote coexistence. Furthermore, these field studies frequently do not examine directly the resources used (types of seed or types of insect) but instead use surrogate measurements (bill size or foraging area). An example of this type of study is Robert MacArthur's work on New England warblers.

Five species of warbler (myrtle, black-throated green, blackburnian, bay-breasted, and Cape May) inhabit the spruce forests of Maine and Vermont in the United States. All of these species nest in the forests of New England and winter in the Caribbean. The spruce forests in which they live and breed are uniform, without obvious variety. Their beaks are all the same size and shape and their stomach contents are approximately the same. In his study, MacArthur recorded how long each species spent foraging in each part of a tree. A tree was divided into 10-ft zones vertically and branches were divided into three horizontal zones: (1) the bare or lichen-covered base, near the trunk, (2) the middle zone of needles, and (3) the terminal zone of new needles or buds. He produced diagrams of trees showing the zones of foraging activity for each species of warbler (Fig. 14). The resulting foraging distributions suggested a rather subtle kind of resource partitioning (assuming that foraging in different parts of the tree gives access to different sections of the insect resource being used by these birds).

In addition to these zone differences, MacArthur also suggested differences in foraging behavior between the five species: "subjectively, the black-throated green appeared nervous, the bay-breasted slow and deliberate." He quantified this by recording the length of time be-

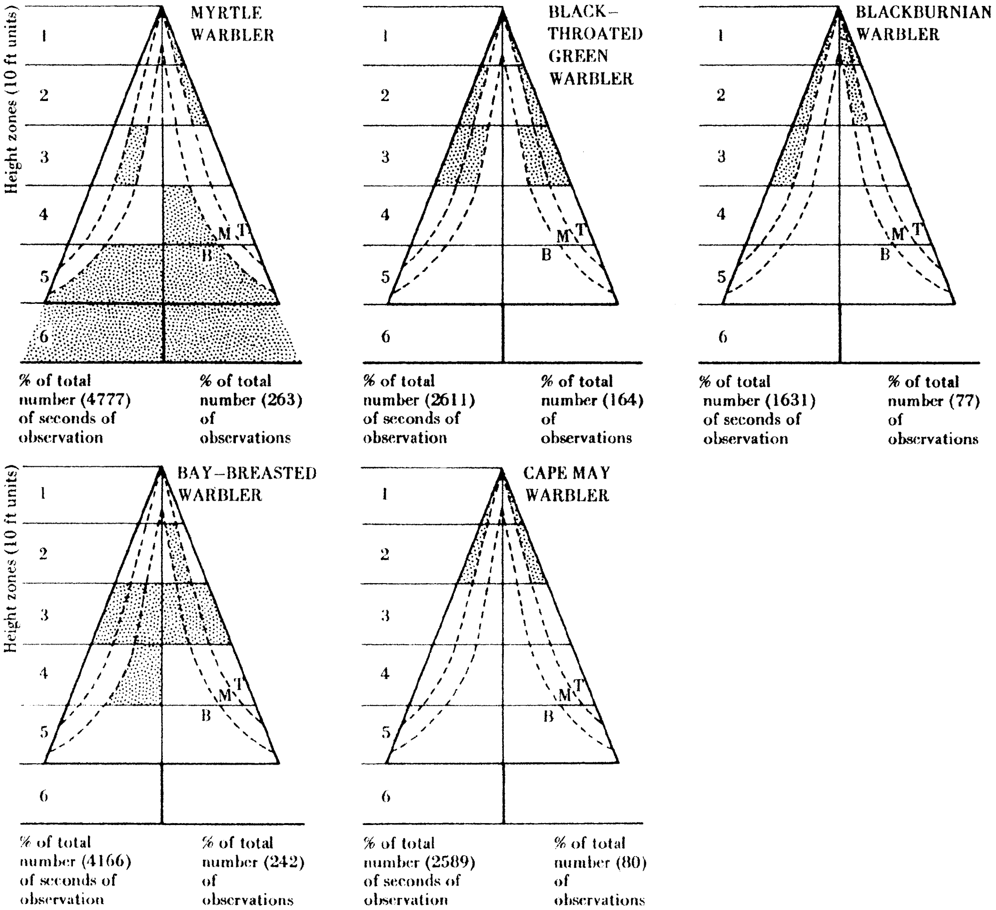


FIGURE 14 Foraging positions of five species of warbler in the coniferous forests of the northeastern United States. Shaded zones represent 50% of recorded time.

tween "use of wings" and was able to show that the black-throated green had a shorter time interval than either the blackburnian or myrtle, which in turn had a shorter time interval than the Cape May and bay-breasted. He also recorded the direction in which each species predominantly moved while foraging. This could be vertical (up and down the tree), tangential (around the tree), or radial (from the center of the tree outward and vice versa). Again he found differences between the species.

MacArthur's work suggests that even when two species overlap in foraging space, they examine that space differently, and therefore use a different part of the available resource. The five warbler species are clearly searching for their insect resources in different spaces and in different ways, but whether this is sufficient to promote coexistence between potential competitors is difficult to tell. In fact, we have no direct evidence that these warblers are even competing for limited resources.

B. Pacific Island Birds

A number of field studies have used patterns in spatial distribution as evidence of the effects of competition on species co-occurrence. This type of field observation is illustrated by Jared Diamond's study (1975) of land-breeding birds on the Bismarck Islands off the coast of Papua New Guinea. Several pairs of ecologically similar species (good candidates for potentially strong competitors) showed what Diamond called checkerboard distributions, for example, the two flycatchers *Pachycephala pectoralis* and *P. melanura dahli* (Fig. 15). These two species have mutually exclusive but interdigitating distributions, with only one of the two flycatchers present on any one island (some islands have neither species).

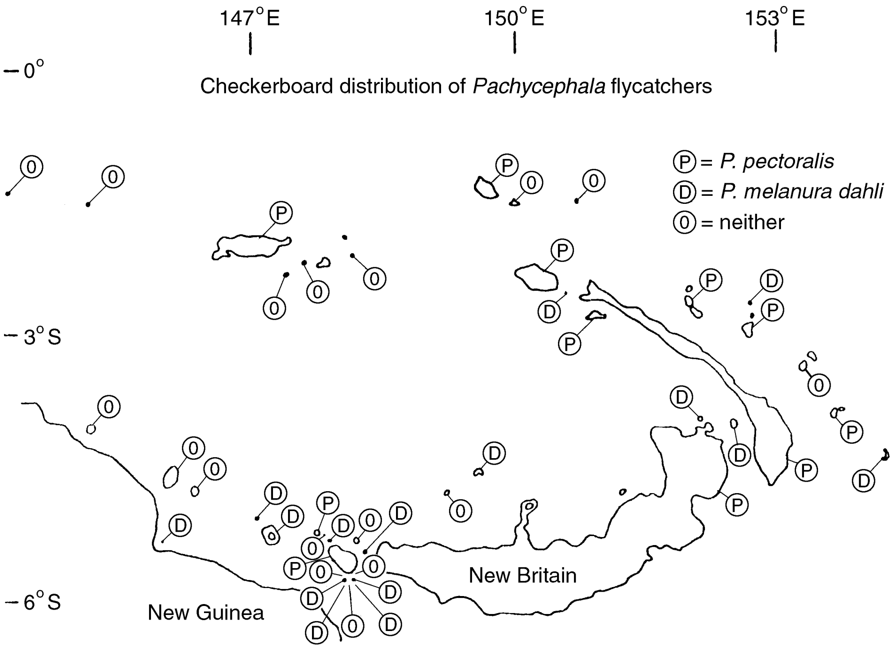


FIGURE 15 Checkerboard distribution of *Pachycephala* flycatcher species on the Bismarck Islands off Papua New Guinea. *P* = *P. pectoralis*, *D* = *P. melanura dahli*, and 0 = neither.

These checkerboard distributions are consistent with the idea that two ecologically similar species are excluding each other from islands because of strong competitive interactions.

Although Diamond did not find many examples of checkerboard distributions among his island birds, he did find other evidence of a more diffuse type of competition by plotting what he called incidence functions. These are graphs in which the probability of occurrence on an island [*J* = (# of islands with the species present)/(total # of islands)] is plotted against the number of species (*S*) on that island. Because of the species–area effect, *S* is also a measure of island size. Figure 16 shows the incidence function for the cuckoo-dove *Macropygia mackinlayi*. This is an example of a species that Diamond calls a "supertramp." These species are excellent colonizers but are poor at persisting in the diverse communities of large islands. This absence of supertramps from large islands implies strongly that competition is important in structuring these island bird communities.

C. African Wild Dogs and Hyenas

Many field observations of "competition" have relied heavily on changing species numbers over time. For example, one species arrives in an area (or a resident species increases its numbers), while another species declines in abundance. An early case was that of two species of crayfish in Russian streams, first commented upon by Gause in his book that reported his early *Paramecium* experiments (Gause, 1934). Another well-cited example is the invasion and spread of the American gray squirrel (*Sciurus carolinensis*) in Great Britain, followed by the decline and contraction of the native red squirrel (*S. vulgaris*). Yet another example is provided

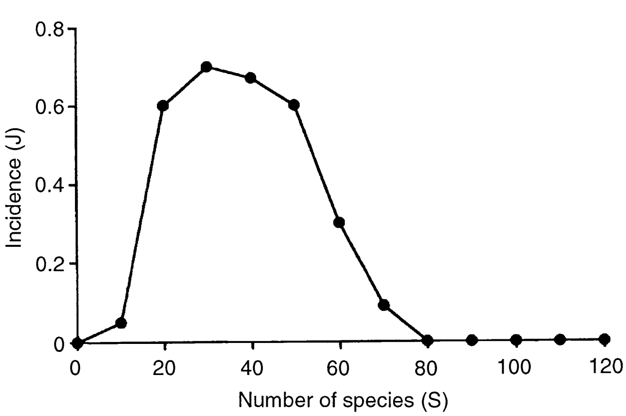


FIGURE 16 Incidence function for the cuckoo-dove *Macropygia mackinlayi* on the Bismarck Islands.

TABLE IX

Population Estimates for the Spotted Hyena Population of the Serengeti

Year	Population estimate
1967	2207 ± 120
1977	3306 ± 432
1986	5214 ± 828
1991[a]	9500

[a] Estimate includes hyenas of the Mara.

by the increase in abundance of the spotted hyena (*Crocuta crocuta*) in the Serengeti-Mara ecosystem of Kenya and Tanzania and the subsequent decline of the African wild dog (*Lycaon pictus*).

Spotted hyenas are the most abundant large predator in the Serengeti-Mara savanna ecosystem. Between 1969 and 1976, the hyena population increased by 50%, probably as a response to the herbivore increase following the elimination of the rinderpest virus. Some estimates for the subsequent population of hyenas in the Serengeti are shown in Table IX. Over this period the wild dog showed a decline in numbers, with all remaining dog packs finally disappearing from the Serengeti in the early 1990s (Fig. 17). During the early 1970s, part of this decline in wild dog numbers was certainly due to hunting, as wild dogs were regarded as vicious vermin to be exterminated. However, the reason for the later decline is less certain, but competition with hyenas may have been important.

Both of these predators hunt in packs and run down their prey (mainly Thomson's gazelle, wildebeest, and zebra). Hyenas frequently follow dog packs when they go hunting from a den (during the breeding season) and steal their kill. Dogs can defend a kill, but about four dogs are required to keep off one hyena. Although food on the "hoof" is probably not a limiting resource for large carnivores in the Serengeti, food in the form of "kills" may be. Particularly during the breeding season, dog packs are hunting not only for themselves but also for the dominant bitch and her pups back at the den. Losing kills to competitors such as hyenas may have a serious effect on pup survival. Because the ability of wild dogs to defend kills depends on the ratio of dogs to hyena, this effect may have become gradually more serious over the last 30 years. Additional evidence supporting this story comes from the correlation ($r = -0.92$, $P = 0.01$) between hyena density and dog density over a series of ecosystems in eastern and southern Africa (Fig. 18).

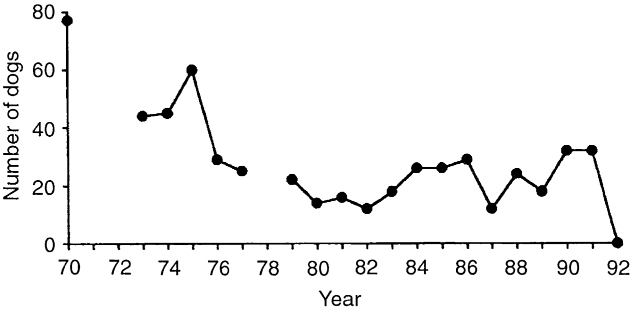


FIGURE 17 Number of wild dogs in the Serengeti ecosystem.

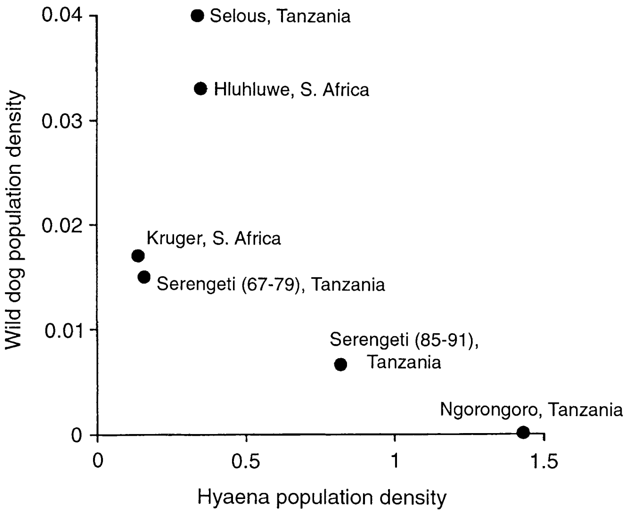


FIGURE 18 Correlation between the population densities of wild dogs and hyenas over a number of African ecosystems.

VI. CONCLUSIONS

The Lotka–Volterra model can theoretically predict the outcome of interspecific competition between two species. Depending on initial population size (N_1 and N_2), carrying capacity (K_1 and K_2), and competition coefficient (α_{12} and α_{21}), either species can be the sole survivor, one species is always the sole survivor (competitive exclusion), or the two species will coexist. What is more, coexistence is assured if $\alpha_{12}/K_1 > 1/K_2$ and $\alpha_{21}/K_2 > 1/K_1$, which implies that intra-specific competition must be greater than inter-specific competition for both species. These predicted outcomes are "local" outcomes (within a habitat patch such as a fungal fruiting body, a freshwater pool, a coniferous wood, or an island depending on the scale of the environmental grain appropriate to the ecology of the organism). Both the aggregation model of competition and resource partitioning predict "local" competitive exclusion, with global coexistence using different local mechanisms that ultimately have the same effect on competitive interactions.

There is considerable evidence, from laboratory experiments, that these models could describe what happens to some real organisms. Graphs of N_1 against N_2 (using K and α) predict the correct competitive outcome for *Paramecium aurelia* and *P. caudatum*, while a small amount of environmental heterogeneity allows sufficient resource partitioning for *P. bursaria* to coexist with *P. aurelia*. The *Tribolium* experiments show that initial numbers can influence which species will win the competitive battle, and these same experiments, plus those with *Drosophila*, show that more complex relationships can still be incorporated into the same theoretical framework.

Lastly, there is some evidence, from field experiments and field observations, that these mechanisms might operate outside the computer and laboratory. We see patterns of replacement reminiscent of competitive exclusion (Bismarck birds and Serengeti carnivores) and niche differences consistent with resource partitioning (New England warblers). We also see numerical changes, after removal or introduction, that are consistent with competitive effects of one species on another (desert ants and mice, Scottish barnacles, and Caribbean lizards). However, the field data tend to fit the patterns predicted by theory rather than providing detailed parameter values for the models. Nonetheless, the ideas encapsulated within theoretical models, such as the Lotka–Volterra model, provide a believable framework within which to interpret complex species interactions in the real world.

See Also the Following Articles

CARRYING CAPACITY, CONCEPT OF • COEVOLUTION • PREDATORS, ECOLOGICAL ROLE OF • RESOURCE PARTITIONING • SPECIES COEXISTENCE

Bibliography

Atkinson, W. D., and Shorrocks, B. (1981). Competition on a divided and ephemeral resource: A simulation model. *J. Animal Ecol.* **50**, 461–471.

Diamond, J. M. (1975). Assembly of species communities. In *Ecology and Evolution of Communities* (M. L. Cody and J. M. Diamond, eds.), pp. 342–444. Belknap, Cambridge, Massachusetts.

Gause, G. F. (1934). *The Struggle for Existence.* Williams and Wilkins, Baltimore.

Gilpin, M. E., and Ayala, F. J. (1973). Global models of growth and competition. *Proc. Nat. Acad. Sci.* **70**, 3590–3593.

MacArthur, R. H. (1972). *Geographical Ecology*. Harper & Row, New York.

Shorrocks, B., and Rosewell, J. (1987). Spatial patchiness and community structure: Coexistence an guild size of drosophilids on ephemeral resources. In *Organisation of Communities: Past and Present* (J. H. R. Gee and P. S. Giller, eds.), pp. 29–51. Blackwell Scientific, Oxford, United Kingdom.

Shorrocks, B., Rosewell, J., Edwards, K., and Atkinson, W. D. (1984). Interspecific competition is not a major organising force in many insect communities. *Nature* **310**, 310–312.

COMPLEMENTARITY

Paul Williams
The Natural History Museum, London

GLOSSARY

efficiency The degree of success in reaching a conservation goal relative to the cost (or to some surrogate for cost, such as the number of areas). For goals such as maximizing the number of species represented for a particular cost, a distinction should be made between merely selecting areas on the basis of any records of species and choosing areas that maximize persistence of species in the long term.

flexibility The degree to which alternatives exist for one or more selected areas in the context of reaching a particular conservation goal. When seeking a set of areas to represent maximum diversity, flexibility for a selected area may be absent, incomplete (replacing the selected area while still reaching the conservation goal would require substitution of two or more areas or one or more areas of greater cost), or complete (other areas could be substituted, one-for-one by number or by cost, with the current choice).

goals An explicit and precise statement of conservation aims. Goals should express the values of those people who provide the mandate for conservation. Making the goal explicit allows efficiency to be measured as an aid to accountability. A statement should include which attributes are valued (such as genetic diversity), which surrogates for this value are actually surveyed (such as higher taxa, species, and threatened species), which areas are to be considered (such as land management units or grid cells), and how constraints of viability, threat, and cost are to be measured and accommodated. Goals are not universal but depend on people's values and their situations. Consequently, different goals may conflict, and areas necessary to meet one goal may be insufficient to meet broader goals.

irreplaceability A property of areas that include species (or other valued attributes) restricted to so few areas that all such areas would be needed in order to meet a conservation goal.

priority Areas that need the most urgent management action to avert threat in order to meet a conservation goal (such as to increase the probability of persistence of valued species or other attributes). Ranking of areas by priority may differ from their ranking by value.

representation The occurrence of species (or other attributes) within a set of selected areas. A distinction must be made between records of a species and areas with high probability of persistence for the species in the long term.

values Values are interpreted here in the broadest sense to include monetary and nonmonetary values. Biodiversity attracts many different values, which are not universal and which depend on social and economic situations. Consequently, different values may conflict.

COMPLEMENTARITY is a property of sets of objects that exists when at least some of the objects in one set differ from the objects in another set. This definition may be extended to include parts of a pattern, such as may be used to predict the distribution of the individual objects. It can be applied to genes, species, species assemblages, or any other attributes of the biota of areas, but it is not limited to these cases. Complementarity is closely related to the idea of spatial turnover (beta and gamma diversity) in the composition of biotas, and it arises because of differences in ecology and history among populations of organisms. For example, two biotas of the same size from areas with grassland and forest may be expected to have a higher proportion of different and complementary species (and therefore will represent more species in total) than would two biotas of the same size from different areas within the same kind of grassland. This article describes complementarity primarily in the context of its application to area-selection problems for biodiversity conservation, describing how complementarity can be used to ensure that as much valued biodiversity as possible is represented for conservation within constraints such as viability and cost. It reviews some of the techniques available and how other needs could be integrated.

I. COMPLEMENTARITY AND AREA SELECTION

A simple example of complementarity is illustrated in Fig. 1, which represents the mammal faunas of two areas. These faunas share one species, the lion, but each fauna also has species that are not shared. It is the species in one fauna that are not shared with the other fauna that are said to be "complementary" with respect to the other fauna.

A. Biological Basis of Complementarity

The biological causes of complementarity are the processes that lead to differences among biotas. At the species level, these processes are familiar from differences in ecology and history. In ecology, differences among species in their needs and responses to the environment, including responses to other species, lead to differences in species distributions near equilibrium. Unique events in history, particularly the gain or loss of barriers preventing any approach to large-scale ecological equilibrium, also contribute to these distributions, as studied in vicariance biogeography (see Section I,E).

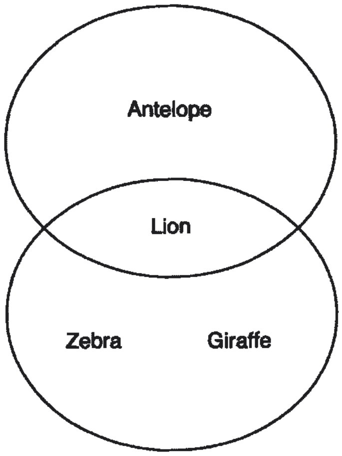


FIGURE 1 Complementarity. The two ellipses represent two areas, one with a fauna of two and the other with a fauna of three mammal species. They share one species, the lion. The zebra and giraffe in the lower area are complementary to the fauna of the upper area, whereas the antelope of the upper area is complementary to the fauna of the lower area.

The description of patterns of complementarity at the species level is closely related to the description of spatial turnover, also known as beta or gamma diversity depending on the spatial scale being studied. There is a substantial literature on spatial turnover, documenting and exploring patterns of species differences, particularly along transects of environmental gradients.

Patterns of complementarity in biodiversity may be expected to be highly dependent on spatial scale because the size of area patches used to study patterns in the distribution of species (for example) has a strong effect on measures of their co-occupancy of these patches. For example, it has been shown that patterns in species richness may differ strongly across the same region when measured using different patch sizes (such as with different grid-cell sizes). Because complementarity also depends on the pattern of species co-occupancy of area patches, the pattern of complementarity may be strongly dependent on the spatial scale at which it is measured.

B. Need for Complementarity in Area Selection

Complementarity is an apparently simple idea that has considerable power to aid area-selection methods for conserving as much biodiversity as possible. Although decisions on area selection will always benefit from experienced judgment, quantitative methods provide a framework of rules designed to make such complex decisions more rigorous and efficient and more open to public accountability. Other properties of these methods include their ability to identify flexibility for the planning process; speed to facilitate the exploration of alternative values, goals, data, and flexible solutions; their ability to deal with incomplete data; and simplicity to aid communication.

Efficiency is important because the area of land (or volume of sea and air) available for conservation is limited since there is often competition between conservation and other, incompatible land uses. However, efficiency should be considered in relation to the goal of ensuring effective long-term persistence of biodiversity. Efficiency comes from using complementarity to choose areas that encompass the most biodiversity within these constraints.

Accountability is important because conservationists should be acting in the broader interests of society, and therefore people should be able to see that their values are being acted on and that limited resources are not being wasted. This could become increasingly important as competition for limited land and resources becomes more intense and as conservationists are obliged to defend their choice of areas. Accountability comes from using quantitative methods in which rules are applied in a consistent manner, forcing people to be explicit about how they might resolve the often conflicting needs of large numbers of species. This makes it possible to trace precisely why one area is chosen in preference to another and, if necessary, to challenge the data or rules.

Complementarity may be particularly helpful in considering options when different land uses conflict because it can identify the unique and valued contribution of the biota in each area. However, the potential of complementarity-based methods will only be fully realized, as with all other area-selection methods, when important constraints such as viability and threat can be assessed realistically.

C. History of Complementarity in Area Selection

Complementarity-based methods were first described by Kirkpatrick (1983). The term "complementarity" was coined specifically in the context of area selection by Vane-Wright *et al.* (1991). The importance of the term is that it identifies a property essential to a group of optimizing techniques (and to more approximate heuristics), which has been useful in comparative studies with other approaches to area selection.

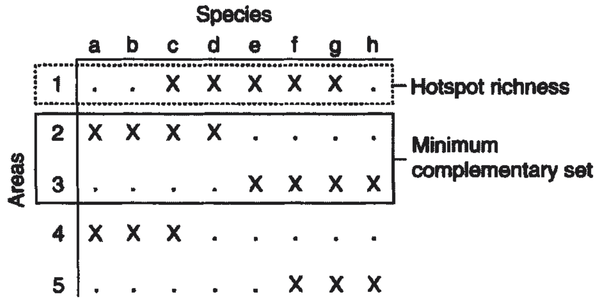
The impetus for the development of complementarity methods in conservation has been strengthened in the past 10 years by Agenda 21 and by the Biodiversity Convention. These policy initiatives gave a new emphasis to conserving the value of broader biodiversity in addition to the long-recognized need to conserve the relatively few, particularly threatened or vulnerable species. Conserving broader biodiversity presented several new challenges, particularly in keeping track of the very large numbers of components of which biodiversity is comprised.

Complementarity-based methods have been used more in the Southern Hemisphere (particularly in Australia and South Africa), where knowledge of the biota often derives from new surveys, some designed specifically for identifying potential conservation areas. In contrast, complementarity methods have been less popular in countries with a long natural history tradition and with a strong local knowledge base, where many locally important areas may already be well-known (although rigorous prioritizing among them or adding to them efficiently may still be difficult without using complementarity). Nonetheless, there is a growing interest in complementarity methods by the U.S. Gap Analysis Program, The Nature Conservancy, Conservation International, and biological recording schemes in Europe. One of the advantages of heuristic techniques for seeking complementary areas is that they are very simple so that they can be implemented easily, which is making them increasingly popular in biodiversity conservation research.

D. Use of Complementarity in Area Selection

Representing as much biodiversity as possible for conservation, given limited resources, essentially depends on the principle of complementarity. To simplify discussion, this article concentrates on the most popular surrogates for biodiversity value—the different species (rather than on genes or ecosystems).

For example, suppose we could choose only two areas from Fig. 2, and we had to do it to represent as much diversity as possible. It might be tempting to begin by choosing area 1 because it has the largest number of species (five). The next richest area is area 2. Areas 1 and 2 represent a total of seven species (a–g).



Areas	a	b	c	d	e	f	g	h	
1	.	.	X	X	X	X	X	.	Hotspot richness
2	X	X	X	X	.	.	.	.	Minimum complementary set
3	.	.	.	.	X	X	X	X	
4	X	X	X	.	.	.	.	.	
5	.	.	.	.	.	X	X	X	

FIGURE 2 Hot spots and complementarity. Five areas have a biota of eight species. Area 1 is the hot spot of highest species richness (five species). However, this area is not part of the minimum fully representative set for all eight species, which is the combination of areas 2 and 3.

However, it can be seen from Fig. 2 that the biotas of areas 2 and 3 are completely different from one another, and if chosen together they would represent all eight species (a–h).

In this example, complementarity maximizes the value represented by area selection, but this is not necessarily the same as identifying priorities. "Priority" is used here for the relative urgency of conservation action among different areas, and this is usually determined from the known threats. For example, in some cases the threat may be to the entire biota of an area, for example, by clear-felling and building. In Fig. 2, even if there were more species and more value in area 3, if the species in area 2 were predictably more threatened then area 2 might have higher priority (see Section III,D).

With all questions of area selection, it is important to specify the goals of the assessment precisely in order to find the most appropriate answers. Even slightly different questions may have quite different answers, and these may be in conflict.

E. Complementarity and Indicators of Biodiversity

Data are rarely available for all the valued parts of biodiversity. One attempt to resolve this problem is to use data for one group as a surrogate or indicator for another or for all biodiversity. This relies on congruence in distributions between the indicator and indicated groups, not just in patterns of richness but also in patterns of complementarity. Whether congruence in complementarity between groups is realistic as a general principle has been questioned for many years, and empirical studies show substantial variation.

Patterns of complementarity might be expected to differ among groups of organisms because the ecological and historical processes that have shaped their distributions differ. In cases in which these processes have been more similar, then a closer correspondence in patterns of complementarity may be expected. For example, some groups share closer ecological similarities in diet, habitat, and climatic tolerances. If some of them have diversified at the same time and in the same areas, they may also have shared similar phylogenetic and biogeographic patterns, as envisaged by vicariance biogeography. When seeking groups of species to use as indicators for use in area selection to represent other groups or broader biodiversity, shared patterns of complementarity within ecological or phylogenetic patterns are needed (Williams *et al.*, 2000).

II. REPRESENTING BIODIVERSITY FOR CONSERVATION

A. Identifying Goals

A common goal of biodiversity conservation projects is to ensure the persistence of as much biodiversity as possible for the future. The basic form of the problem is one of optimizing the number of valued components of biodiversity (usually species) represented within a set of areas (whether these are to be protected or not), subject to constraints including viability, threat, and cost. These constraints have often been considered at a subsequent stage, in which the component areas of a proposed protected area network are refined, usually by moving to finer spatial scales. Ideally, however, all constraints could be included in the overall design of a network, if the candidate areas were realistic land management units and if the data were available. In principle, this could include appropriate treatments of ecological processes and climate change.

A simple representation goal used to illustrate the basic principles in many academic studies has been to achieve at least one representation of every included species within the reserve network. Alternatively, the goal could be to achieve any required number of representations (this number could also differ among species) or could be expressed in terms of population size, probability of occurrence, etc. For many studies to date, information on constraints has been less widely quantified (although perhaps because less effort was invested).

Viability and threat are often either ignored or treated only by using crude rules (see Section II,E). The cost constraint is usually represented by the number of area units required or sometimes by the area extent (see Section II,D).

One way of dealing with threat has been to include only the most threatened species for area selection (see Section III,D). This may amount to "fire fighting" (the reactive approach) as and when species become threatened, which may risk concentrating resources in the "wrong" place in the long term if threat changes unpredictably. The alternative proactive or strategic approach is to include all species for representation, which requires investing resources in at least surveying all the species, including those that are not currently threatened. The proactive approach does not assume that simple protection of one area is an adequate or appropriate action for all species. Representation goals may still differ among species and should reflect their different needs, although this would require a rigorous justification to make the process defensible.

An important part of specifying an area-selection goal may be to include the contribution of an existing conservation network. If me species can be considered to be adequately protected by the existing network, then only the complementary species need be targeted for additional representation. This is the idea behind "gap analysis," which can be implemented using complementarity-based procedures.

B. Minimum-Cost Sets

One kind of question is of the following form: "Which is the minimum set of areas within Madagascar required to represent the diversity of butterflies?"—to achieve at least one representation (or any other required number) for species within the area set. This "minimum set" form of the optimization problem has been stated formally as an integer-programming problem by Camm *et al.* (1996):

$$\min \sum_{\{j \in J\}} x_j$$

$$\begin{array}{lll} \text{subject to} & (1) & \sum_{\{j \in Ni\}} x_j \geq 1, \text{ for all } i \in I \\ & (2) & x_j = (0, 1), \text{ for all } j \in J \end{array}$$

where $J = \{j | j = 1 \ldots n\}$ is the set of candidate areas for selection, and $I = \{i | i = 1 \ldots m\}$ is the set of species to be represented. N_j is the subset of areas J that contain species i. The selection variable $x_j = 1$ if the area j is selected or 0 if it is not selected. The first constraint is that every species must be represented in at least one of the areas (or more areas for multiple representation goals), with one such constraint for each of the m species. The second constraint states that selection is all or nothing for each of the n areas.

Minimum sets aim to achieve complete representation, usually in the first instance by seeking representation of every species in at least one area (Fig. 3) or in more areas for multiple representation goals. There are many pitfalls with this approach, not only concerning the problems of viability and threat that are discussed later but also concerning the meaning of "complete" in this context. It is apparent that a single representatn of a species will not necessarily represent all of the intraspecific variation. The problem becomes worse when more remote surrogates for biodiversity value are used. For example, when using ecoregion or community data, such as vegetation classifications, a single representation of every vegetation class would be very unlikely to represent every species, let alone all intraspecific variation. Of course, the only reliable solution to representing every difference is to include every area, although the premise of using area-selection methods is that competition with incompatible land uses limits the area available for conservation. Nonetheless, minimum sets for species at least have the advantage that they are expected to represent more species than any undirected selection for the same number of areas or cost.

C. Maximum-Coverage Sets

Another kind of question takes the following form: "How can we choose 1% of the total area of Madagascar to represent the greatest diversity of butterflies?" This "maximum-coverage" form of the optimization problem has been stated formally as an integer-programming problem by Camm *et al.* (1996):

$$\max \sum_{\{i \in I\}} y_i$$

$$\begin{array}{lll} \text{subject to} & (1) & \sum_{\{j \in Ni\}} x_j \geq y_i, \text{ for all } i \in I \\ & (2) & \sum_{\{j \in J\}} x_j \geq k \\ & (3) & y_i = (0, 1), \text{ for all } i \in I \\ & (4) & x_j = (0, 1), \text{ for all } j \in J \end{array}$$

where I, J, N_i, and x_j are defined as for the minimum set problem. In addition, constraint (2) is the number

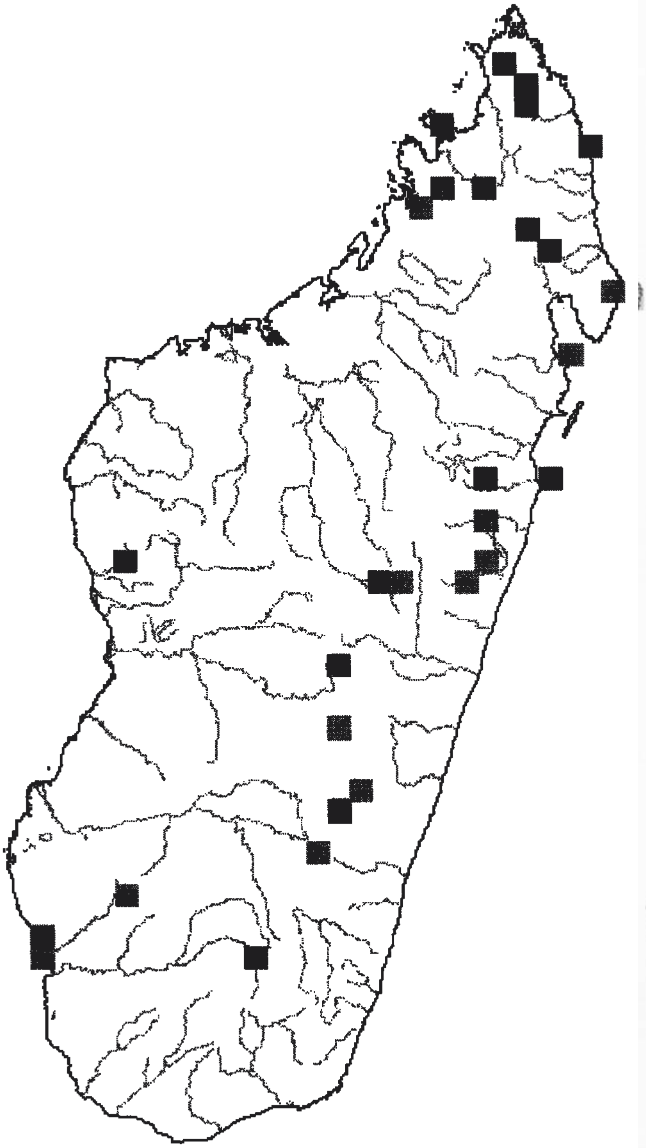


FIGURE 3 Near-minimum cost set. Example of 29 quarter-degree grid cells (each approximately 170 km²) selected using a heuristic algorithm to represent all of a group of 321 species of Madagascan butterflies (ignoring other constraints and using the number of cells as a surrogate for cost). The black areas are irreplaceable, whereas the gray areas are flexible because there are alternatives to these areas. The numbers show how many species each area represents uniquely within this set (data courtesy of David Lees).

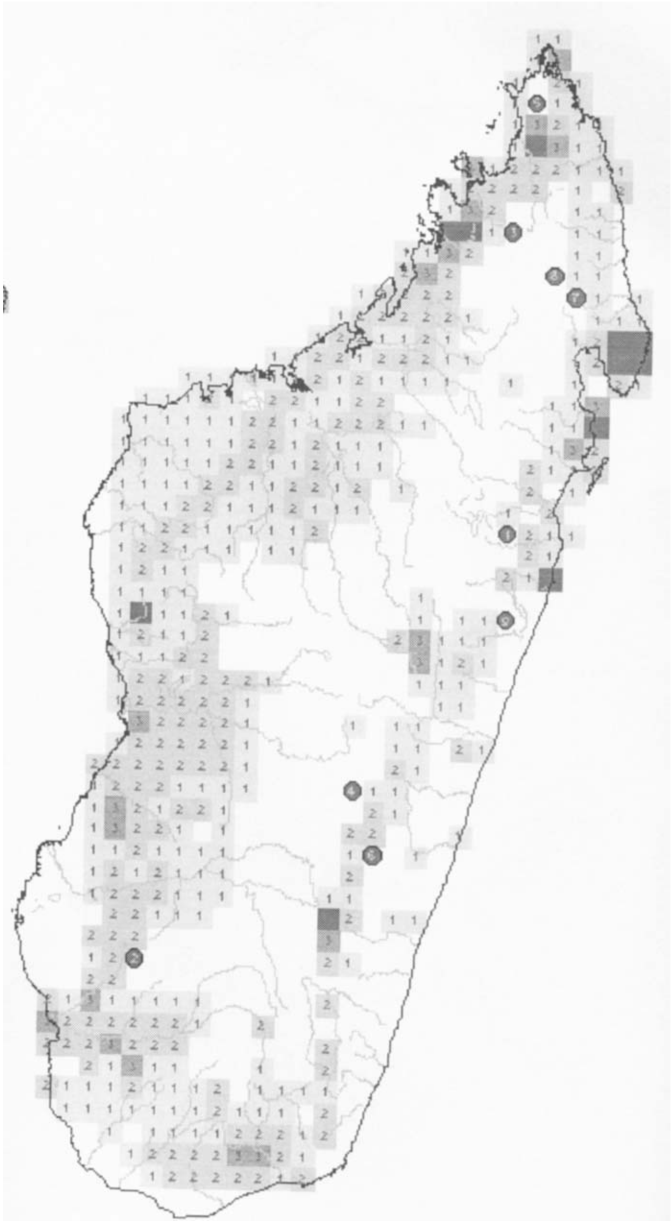


FIGURE 4 Near-maximum coverage set. Example of 1% of quarter-degree grid cells (nine cells) selected using a heuristic algorithm to represent as many as possible (87%) of a group of 321 species of Madagascan butterflies (ignoring other constraints). The gray spots with numbers show the nine selected areas in order of complementary species richness. The gray squares with numbers show richness in the 41 species of butterflies not represented within the nine selected areas (data courtesy of David Lees).

of areas to be selected, *k*. Because not all species will be represented, the constraints are modified, with $y_i = 1$ if species *i* is represented in at least one selected area or 0 if it is not represented.

Maximum-coverage problems (Fig. 4) may often prove to be the more realistic and widely applicable form of the problem because conservation goals are usually constrained more by budget and competing land uses than by ideas of complete representation. It is therefore unavoidable to have to face the "agony of choice" that not all valued biodiversity can be represented in areas selected for conservation (Vane-Wright *et al.*, 1991).

D. The Cost Constraint

In the simplest formulation of conservation goals, the cost constraint on area selection can be represented by the number of areas, treating all areas as being of equal cost. A few studies have used the extent of areas of differing size as a measure of cost. Even when financial cost data have been used, they have rarely included a thorough consideration of all the appropriate cost components. These may include (but are not limited to) cost of area acquisition, costs of management to reduce threats, and (crucially) the opportunity costs that arise from income from any alternative uses of an area (e.g., forestry and agriculture) that is foregone when it is managed for biodiversity conservation (Faith and Walker, 1996).

When information on how cost varies among areas is available, it can be included in area selection by dividing the complementary part of the biodiversity value of areas by the cost of those areas to calculate the cost of acquiring a particular biodiversity increment.

The consequence of including cost data that differentiate among areas when seeking a minimum representative set is usually to decrease the cost, although the number of areas or the area required usually increase. For goals of finding a maximum-coverage set for a given cost, the number of areas and the amount of biodiversity represented usually both increase. If, however, sets are sought for increasing numbers of areas so as to maximize the numbers of species at each step, then the minimum cost may fluctuate strongly as the number of areas increases (because it may be cheaper to acquire more areas for representing the same number of species).

E. Other Constraints

Ultimately, effective biodiversity conservation is about ensuring the persistence of biodiversity values. This requires that appropriate treatment of viability and threat be integrated into area-selection methods. With branch-and-bound techniques (see Section III,B), multiple constraints can be included in the specification of the problem. With heuristics (see Section III,C), a variety of options for integration have been used. Those that are structured around a sequence of decisions (as opposed to combining or weighting scores for the different constraints) can have the advantage of better accountability because accounting for why an area is included or excluded at each stage is straightforward (Williams, 1998). For example, constraints may be imposed by applying a sequence of filters to the data (Bedward *et al.*, 1992). To deal with viability, the simplest treatment is to include only those populations likely to be viable, although the necessary data to make this determination are rarely available. To deal with threat, areas in which species are known to be under severe threat might be excluded from consideration because they are deemed to be irretrievable, or the cost of averting threats might be included in the area cost.

III. COMPLEMENTARITY-BASED TECHNIQUES

In addition to the biological and social problems of accounting for viability and threat, area-selection methods have to deal with a technical problem of how to represent the most biodiversity (as species or any other valued aspects of biodiversity) with limited resources. Achieving the representation goals within the constraints described in the previous section is not entirely straightforward. For example, the size of the problem when seeking a minimum-cost set of areas cannot be calculated simply from the amount of data (this problem is described as not polynomial complete or "n-p complete"). Fortunately, the problem is familiar to mathematicians from many kinds of similar applications in operations research, in which a range of techniques for dealing with it have gained acceptance (exhaustive search, branch-and-bound, and heuristics).

The size of the area-selection problem may often be reduced before applying any of the techniques described later by simply removing redundant data. For example, if two areas have exactly the same species, then one of the areas can be removed from the data, providing a record is made that the data for the retained area represent either of the original areas. Similarly, if two species have the same distribution, then one species may be removed. This may help to speed up the procedure, but it is only applicable if the viability and threat treatment for the retained data and the removed data are identical. Similarly, widespread generalist species may not need protection to persist and could be excluded from an analysis.

A. Exhaustive Search

The procedure for exhaustive search is to test every possible combination of areas to find the optimum.

The advantage is that this provides an exact solution, guaranteeing that the one or more truly optimal results will be found. The disadvantage is that it is only applicable to problems with so few areas that it may be of little practical use. This is because the number of possible combinations to be tried becomes too large to be searched exhaustively, even with very powerful computers. For example, for a simple maximum-coverage problem, when seeking five areas to represent the most species from among 1000 possible candidate areas, there are $1000 \times 999 \times 998 \times 997 \times 996$, or nearly 10^{15} possible combinations of areas to be tested.

B. Branch-and-Bound Algorithms

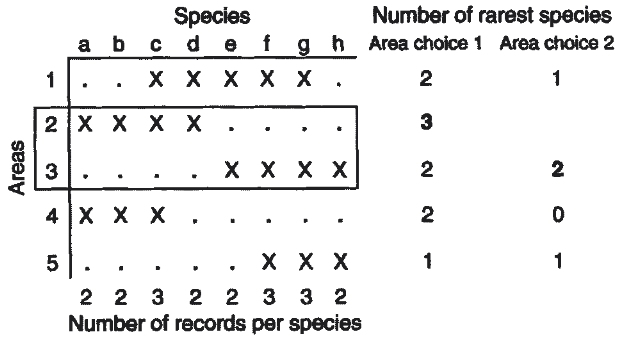
The procedure for branch-and-bound algorithms is to restrict the search to only a subset of all possible combinations of areas, but a subset within which the optimal solution is expected to lie. There are two advantages to branch-and-bound algorithms. First, they can give an exact solution with the truly optimal result. Second, because only a subset of area combinations is searched, this is faster than an exhaustive search and therefore may be applied to larger problems. There are two disadvantages. First, even searching a subset may require very large numbers of combinations to be searched, thus taking a long time. Advances in computer technology have reduced this to days or even hours for large problems, although this may still be too long when the decision-making process demands interactive analysis. Second, when seeking more complex representation goals, for example, involving quantitative area targets, the problem may not be soluble by this approach (Pressey *et al.*, 1996). Branch-and-bound algorithms are also more complicated to implement, although commercial software packages are available.

The branch-and-bound technique begins by setting "bounds" between which the optimal solution must lie. This may be done by solving a simpler form of the problem (a linear-programming relaxation of the integer-programming, or presence–absence, problem), for which each species may be represented a fractional number of times. The technique is then to search "branches," or groups of possible area sets. Any area sets that exceed the bounds can be rejected, along with all of the possible but untested area sets remaining in the branch, thereby greatly reducing the size of the problem. The bounds are usually refined as the search proceeds.

C. Heuristic Algorithms

The procedure for heuristic algorithms is to use relatively simple sets of rules that have been found, from experience with previous data sets, to give good approximations to optimal solutions. There are three advantages to heuristic algorithms. First, although some heuristics are known to be poor, others give results that are usually close to the optimum (Csuti *et al.*, 1997) (hence, heuristics give "near-optimal" solutions) and distinctly better than those obtained from simple scoring techniques that do not use complementarity (Pressey and Nicholls, 1989). In practice, the suboptimality may be trivial compared to concerns about data quality and practical implementation for conservation (Pressey *et al.*, 1996; see Section IV,A). Second, heuristics are by far the fastest complementarity-based techniques available, which makes them suitable for an interactive decision-making process. Third, because they consist of simple rules, they are easy for nonspecialists to implement, with or without a need for computers. There are two disadvantages. First, it is not guaranteed that they will find the optimal solution, only approximations to it. Second, because their relative efficiency depends not only on the choice of rules but also on the particular pattern of species distributions within each data set, it may be worthwhile to apply more than one set of heuristic rules in order to compare the results and select the best (this may still be much faster than a truly optimizing algorithm).

The simplest heuristic procedure is the greedy richness algorithm, which selects the area with highest richness in complementary species (i.e., richest in those not represented at previous steps) at each step or "iteration" of this selection rule until the selection goal is reached. This tends to work well for near-maximum coverage when choosing relatively few areas compared to the minimum set, but it is less efficient as the number of areas approaches the size of the minimum set (Csuti *et al.*, 1997). Area choices for a minimum set are usually most constrained by the more restricted species; therefore, heuristic near-minimum set algorithms that focus on "rarity" (in the sense of the species frequency of occurrence among areas that are candidates for selection) tend to be more efficient. Two kinds of rarity-based algorithms have been used (see Section IV,C). First, weighted rarity algorithms (Kirkpatrick, 1983) are similar to the greedy richness algorithm but weight each species according to its frequency in order to favor choosing areas with many of the more restricted species. Second, progressive rarity algorithms (Margules *et al.*, 1988) identify the rarest complementary species at each step and score areas for richness in only these species, moving progressively to the next less restricted species only after the most restricted species are all represented within selected areas (Fig. 5). Additional rules may be added to deal with ties, and which rule is used to make



Areas	a	b	c	d	e	f	g	h	Number of rarest species Area choice 1	Number of rarest species Area choice 2
1	.	.	X	X	X	X	X	.	2	1
2	X	X	X	X	.	.	.	.	3	
3	.	.	.	.	X	X	X	X	2	2
4	X	X	X	.	.	.	.	.	2	0
5	.	.	.	.	.	X	X	X	1	1
Number of records per species	2	2	3	2	2	3	3	2		

FIGURE 5 Heuristic area-selection algorithm. Example of the application of progressive rarity rules for selecting a near-minimum set of areas for representing all species. First, the numbers of records are counted for each species and the rarest species are identified (those found in two areas). The first choice area is area 2 because it has the largest number of "rarest" species (three: a, b, and d) that are recorded from just two areas. The second choice is area 3 because it has the largest number of complementary (unrepresented) rarest species (e and h). For a goal of representing each species at least once, selection now stops because the goal has been attained. If at the second or subsequent choice there had been no more species with just two records, then the number of next-rarest species, with three records, etc., would have been considered.

TABLE I

Near-Minimum Area Set by a Heuristic Algorithm: Summary of the Progressive Rarity Rules Used at Each Area-Selection Step[a]

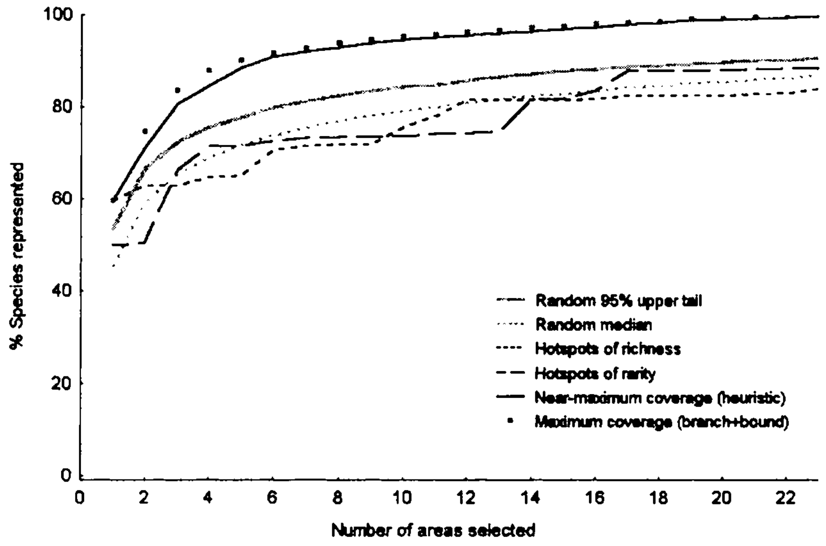
Step	Choices Area	No. of ties	Selection rules, No./ description
1	Kashmir	—	Irreplaceable species
2	Afghanistan	—	Irreplaceable species
3	Ecuador	0	2: rarest + next rarest
4	Michoacan	1	4: rarest + next rarest + next rarest + lowest cell
5	Qinghai	3	4: rarest + (no next rarest) + lowest cell
6	Northeast India	1	4: rarest + (no next rarest) + lowest cell
7	Central Bolivia	0	1: rarest
8	North California	0	3: rarest + next rarest + next rarest
9	Uzbekistan	2	4: rarest + (no next rarest) + lowest cell
10	Turkey	1	4: rarest + next rarest + next rarest + lowest cell
11	Irkutsk	0	2: rarest + next rarest
12	Big Horn	11	4: rarest + (no next rarest) + lowest cell

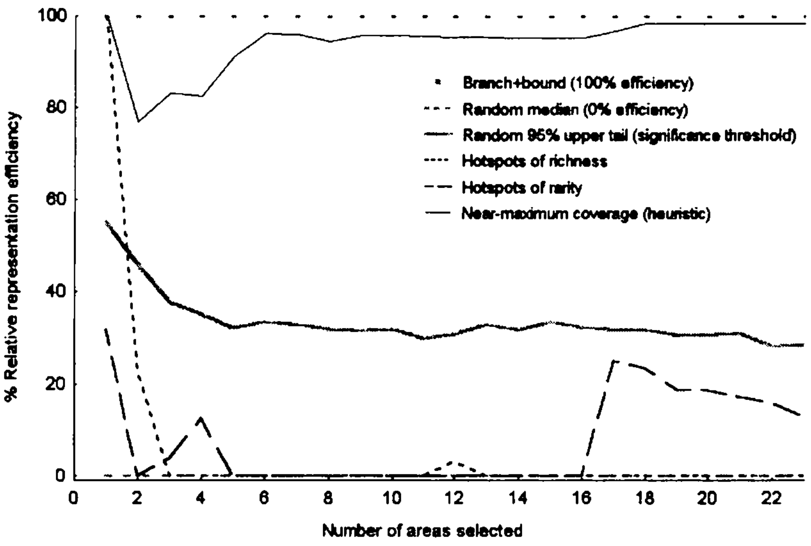
[a] This is a simple example to illustrate the principle of selecting a near-minimum area set to represent all 43 species of bumble bees of the *sibiricus* groups among large equal-area (611,000 km^2) grid cells (Vane-Wright *et al.*, 1991). The table lists the areas in the order (step) in which they have been chosen, together with the rules that were used to choose them. Where a species occurs in only one grid cell, that cell is irreplaceable (areas 1 and 2) if the selection goal is to be met. Once all such species are accounted for, up to four rules are applied (in this case) to make a subsequent choice at each step. The first rule is to select the area with the largest number of the rarest unrepresented species (e.g., for area 7). If there are tied areas for this rule, then the second rule is to choose among the ties in favor of the area with the largest number of the next-rarest unrepresented species (areas 3 and 11). If there are tied areas for this, then the third rule is to choose among them for the area with the largest number of the next-next-rarest unrepresented species (area 8). If ties persist or if there are no next- or next-next-rarest species, then a fourth arbitrary rule of choosing the tied cell with the lowest index number is used (areas 4–6, 9, 10 and 12). Random choices can be used as another option, although the results are unrepeatable for test comparisons. The number of ties shown is the number persisting when the last rule at each step is applied (this is not the same as flexibility).

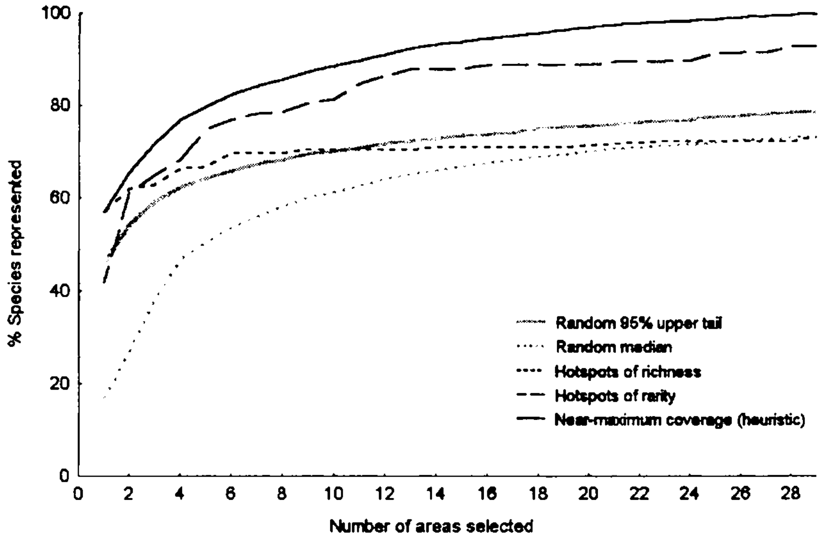
each area choice can be recorded to aid accountability (Table I). This tends to work better for near-maximum coverage when choosing a number of areas that approaches the minimum set, but it is less efficient for very small numbers of areas (Fig. 6a). Which of these heuristic algorithms is most efficient depends on patterns in the data (see Section IV,A).

Heuristic algorithms may be supplemented with additional procedures designed to improve their efficiency. The most basic of these are redundancy tests, which are applied to area sets in order to identify and reject any areas (chosen at early steps of the algorithm) in which all of the species are found subsequently to be represented in other selected areas (chosen at later steps of the algorithm). In extended redundancy tests, the unique contributions of each selected area to the representation goal are identified (see Section IV,B). In some cases, it is then possible to substitute a single area for pairs of areas while still meeting the representation goal.

Re-ordering a near-minimum area set of selected areas by complementary richness may be used as a heuristic technique to provide a series of approximations to maximum-coverage solutions (Table II). Again, the question of which heuristic provides the most efficient result may depend on patterns in the data and on how many areas are needed in comparison with the size of the minimum set. Therefore, if after reordering of the areas the results of cumulative species representation from the different algorithms are compared, then the more efficient maximum-coverage solution for a given number of areas can be chosen. These different area sets must not be mixed subsequently because maxi-


(a) Oregon vertebrates


(c) Oregon vertebrates


(b) Madagascan butterflies

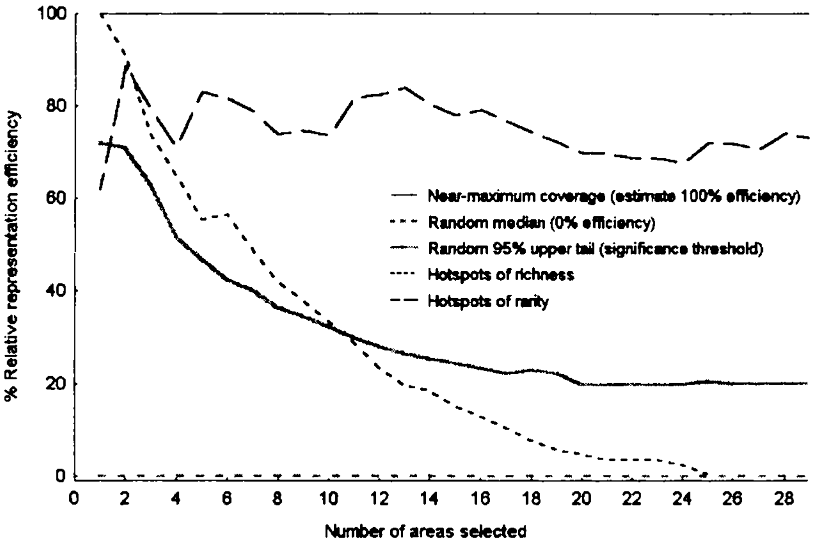

(d) Madagascan butterflies

TABLE II

Near-Minimum Area Set Reordered by Complementary Richness (No Within-Set Redundant Areas Detected): Summary of Species Representation by Area Sequence

Choices		Species richness			
Step	Area	Absolute	Increment	Cumulative	% Total
1	Ecuador	10	10	10	23.26
2	Kashmir	9	9	19	44.19
3	Turkey	7	6	25	58.14
4	Michoacan	4	4	29	67.44
5	Central Bolivia	8	3	32	74.42
6	North California	4	3	35	81.40
7	Irkutsk	5	3	38	88.37
8	Afghanistan	5	1	39	90.70
9	Qinghai	8	1	40	93.02
10	Northeast India	5	1	41	95.35
11	Uzbekistan	4	1	42	97.67
12	Big Horn	4	1	43	100.00

[a] This is an example of the results of reordering the near-minimum set shown in Table I that represents all 43 species of bumble bees of the *sibiricus* group among large equal-area (611,000 km^2) grid cells. The areas are listed by complementary species richness: the order in which they contribute the largest number of previously unrepresented species. The complementary richness at each step or species increments are shown in the fourth column.

mum coverage depends on complementarity among all of the preceding areas within each reordered list. Reordering can also be extended to sets for multiple representation goals (Williams *et al.*, 2000), thereby providing a heuristic approach to near-maximum coverage for any number of areas.

D. Prioritizing Areas

After selection of areas to meet species representation goals, ordering of areas (or reordering when using heuristics) may be used to rank the selected areas by priority. Priority is the urgency with which areas require management action for conservation to avert threats to the persistence of their biodiversity value. Therefore, priority ranking of areas may differ from their ranking by biodiversity value. The distinction between value and priority is important but often ignored.

The choice of criterion for assigning priority may depend on the particular situation. First, priority has been based on measures of threat, but only when threat is predictable and when suitable measures are available,

FIGURE 6 Species representation and relative efficiency. (a) Cumulative percentage of a group of 426 species of Oregon vertebrates represented within increasing numbers of hexagonal grid cells (each 635 km^2, as developed for the Oregon Gap Analysis Program) selected using maximum coverage (branch-and-bound algorithm), near-maximum coverage (progressive rarity algorithm), hot spots of richness, and hot spots of rarity (richness in the rare quartile of species), and areas chosen at random. Scores above the solid gray line (the percentage threshold score to the top 5% within 1000 randomly drawn scores) are significantly better than expected when choosing areas at random (data reanalyzed from Csuti *et al.*, 1997). (b) Cumulative percentage of a group of 321 species of Madagascan butterflies represented within increasing numbers of quarter-degree grid cells (each approximately 170 km^2) selected using near-maximum coverage (progressive rarity algorithm), hot spots of richness, and hot spots of rarity (richness in the rare quartile of species) and among areas chosen at random (data courtesy of David Lees). (c) Results from Fig. 6a rescaled for relative efficiency between the maximum that can be achieved (using the branch-and-bound algorithm) and the median representation expected from selecting areas at random (relative representation efficiency: $E_{\text{observed}, n} = 100(S_{\text{observed}, n} - S_{\text{random}, n})/(S_{\text{maximum}, n} - S_{\text{random}, n})$. Scores above the solid gray line are significantly better than expected when choosing areas at random. (d) Results from Fig. 6b rescaled for relative efficiency between an approximate estimate of the maximum coverage that can be achieved (using a progressive rarity algorithm) and the median representation expected from selecting areas at random. Because the heuristic algorithm is likely to be further from the optimum when very small numbers of areas are selected, the corresponding estimates of relative efficiency may be less reliable.

in order to ensure that the most threatened biodiversity is protected as soon as possible. Second, priority has been based on flexibility to ensure that the parts of biodiversity with fewest options for representation are represented as soon as possible (see Section IV,C). A potential (although unlikely) problem with this option is that unpredictable threats might conceivably still destroy the bulk of biodiversity in areas low on the list, despite their greater flexibility, before it can be protected. Third, when no other information is available, or when threats are intense and unpredictable, a precautionary approach has been to prioritize areas by reordering them in an approximation to a series of maximum-coverage solutions to ensure that the most biodiversity is protected as soon as possible.

IV. ASSESSING COMPLEMENTARITY

A. Efficiency

Much of the emphasis on using complementarity in area selection is a result of the fact that it can be used to represent more biodiversity with limited resources. In general, the more constraints placed upon the optimization problem, the greater the cost. The relationship between representation and cost has been identified with concepts of efficiency. However, efficiency should be considered in relation to not just representation of species records but also in the context of all the other constraints (viability, threat, etc.) necessary to reach the goal of ensuring the effective long-term persistence of biodiversity.

For minimum-set goals, the efficiency of species representation was measured by Pressey and Nicholls (1989) as the proportion of all areas that is not required to meet the goal:

$$E_{\text{minset}} = 1 - (X_{\text{minset}}/T)$$

where E_{minset} is the efficiency, X_{minset} is the number or extent of areas needed for the minimum set, and T is the total number or extent of candidate areas for selection. This is an absolute measure in the sense that efficiency is judged relative to selecting all areas or none.

Because the absolute efficiency that can be achieved in any particular case varies with patterns in the data, it may be more useful to measure efficiency relative to what might actually be expected with those particular data. The upper bound to this is the maximum representation within the chosen constraints, the maximum-coverage set ($s_{\text{max},\,c}$) for the given constraints (c). One useful lower bound for comparison is the representation that might be expected if areas were selected at random, without regard to their biota. In the past, efficiency of species representation has often been assessed informally by comparing species-accumulation curves for ordered sequences of selected areas. Some studies have also used simulated random draws of areas (without replacement) to estimate whether the representation achieved by an area set is significantly greater than would be expected by chance when selecting the same number of areas (Figs. 6a and 6c). This kind of assessment might be refined in many ways, particularly for dealing with existing networks of protected areas. One lower bound of interest might be the median score expected by random draws ($s_{\text{rand},\,c}$) for the same constraints, although other lower bounds could be substituted when appropriate. Therefore, for the representation within an observed set of areas ($S_{\text{obs},\,c}$) for the same constraints, the relative efficiency of any set of areas ($E_{\text{obs},\,c}$) for constraints c is given by

$$E_{\text{obs},c} = (s_{\text{obs},c} - s_{\text{rand},c})/(s_{\text{max},c} - s_{\text{rand},c})$$

In the simplest form of this measure, the number of areas can be used as a measure of the cost constraint, in which case efficiency is measured relative to what might be expected from the data when selecting a particular number of areas (Figs. 6b and 6d).

Higher efficiency in species representation was the original reason why Kirkpatrick (1983) used a complementarity-based technique for area selection rather than simple scoring of areas. This higher efficiency was demonstrated using the absolute measure of efficiency by Pressey and Nicholls (1989). A comparison of complementary areas with hot spots of richness and rarity (Table III) shows that using complementarity results in increased species representation and does so particularly for the most restricted species.

The maximum absolute efficiency in species representation that can be achieved using complementarity is determined by patterns in the data of co-occupancy of areas by species. For example, if all species were found in one area, then only one area would be needed as a minimum to represent them; however, if none of the species were found together, then as many areas as species would be needed (Vane-Wright *et al.*, 1991). In a recent study, absolute efficiency was found to be increased by at least three factors (Pressey *et al.*, 1999). First, efficiency increased with greater "nestedness" among species distributions (perfect nestedness is the pattern in which species present in any particular biota are present in all larger biotas, and where species absent

TABLE III

Representation of a Group of 321 Species of Madagascan Butterflies among 22 (2.5% of the Area) Quarter-Degree (Each Approximately 170 km²) Grid Cells Selected by Three Methods

	Area-selection method		
	Hot spots of richness (%)	Hot spots of rarity (richness in rare quartile of species) (%)	Near-maximum coverage (complementarity) (%)
Number of all (321) species within 2.5% of area	231 (72)	287 (89)	314 (98)
Number of rare-quartile (76) species within 2.5% of area	21 (28)	57 (75)	70 (92)

from any particular biota are absent from all smaller biotas). Second, efficiency increased when species were more broadly distributed. Third, efficiency increased with greater variation in the size (or cost) of areas.

B. Flexibility

Quantitative area-selection techniques frequently identify several alternative combinations of areas that meet their representation goals. This allows for flexibility in the final area choices (Pressey *et al.*, 1993). In general, the more constraints placed on the optimization problem, the less the flexibility. In contrast, areas for which there are no flexible alternatives are "irreplaceable" if the representation goal is to be met.

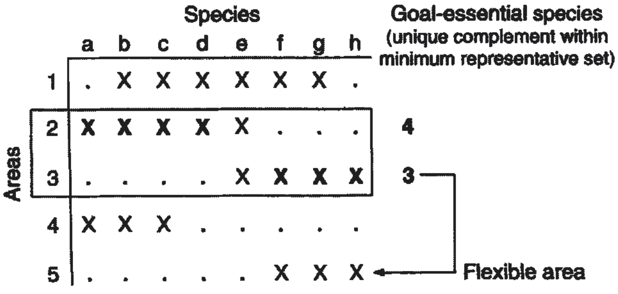
When using exhaustive or branch-and-bound area-selection techniques, flexibility is identified directly as alternative optimal sets. When using heuristic techniques, however, it is done indirectly by identifying which species each selected area contributes uniquely to the representative set (Figs. 3 and 7). Any other area that shares these same "goal-essential" species (Rebelo, 1994) could be substituted in its place (Figs. 7 and 8a) and is therefore a fully flexible alternative (Fig. 8b). Flexibility can also be used to order selected areas by the number of options for representing their goal-essential complement, as shown by the color scale from black to light gray used for selected areas in Fig. 8b.

For truly optimal results, all constraints that can reasonably be quantified should be included when formulating the goal of the original area-selection problem. However, flexibility may be useful for exploring constraints that were not easily quantified or included as part of earlier steps in the area-selection procedure. Decision makers may find interactive exploration of flexibility particularly useful when asking "what if. . .?" questions. Interactive decision making is popular in other fields because it can use the graphical strengths of geographic information system technology to communicate large amounts of information quickly and easily.

C. Irreplaceability

Areas within a representative set for which there are no flexible alternatives (in order to meet a particular goal) have been described as "irreplaceable," an all or nothing property. However, "shades of irreplaceability" has also been used to refer to the relative number of alternatives that may exist for an area choice, depending on how widespread the species occurring there may be (Pressey *et al.*, 1994).

Shades of irreplaceability concerns the frequency with which an area is required among all possible fully



Areas	a	b	c	d	e	f	g	h	Goal-essential species (unique complement within minimum representative set)
1	.	X	X	X	X	X	X	.	
2	X	X	X	X	X	.	.	.	4
3	.	.	.	.	X	X	X	X	3
4	X	X	X	.	.	.	.	.	
5	.	.	.	.	.	X	X	X	

FIGURE 7 Flexibility. Five areas have a biota of eight species. The minimum fully representative set for all eight species is a combination of areas 2 and 3. Area 2 uniquely contributes four species (a–d) to the goal of a single representation of each species, and area 3 uniquely contributes three (f–h) of these goal-essential species. Note that all three of the goal-essential species of area 3 (f–h) are shared by area 5, which is therefore a fully flexible alternative.

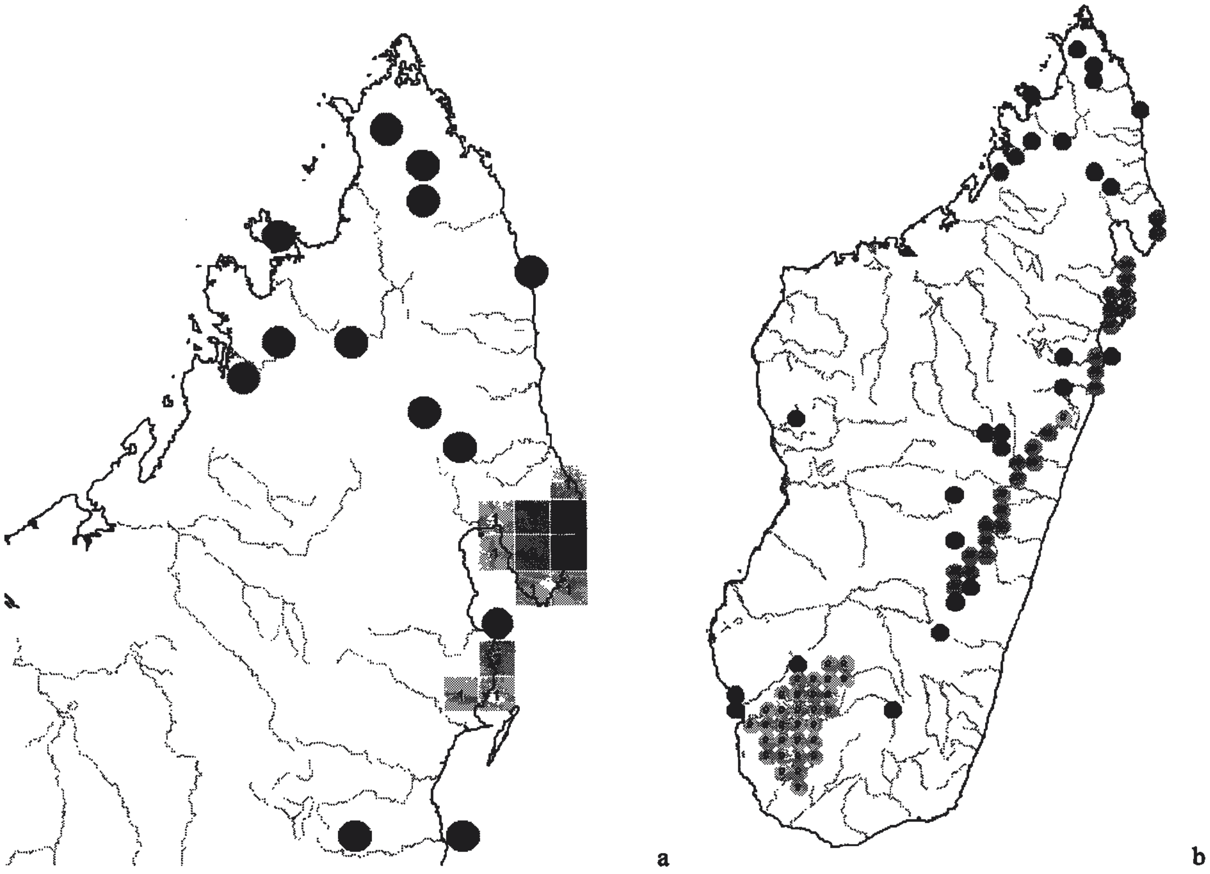


FIGURE 8 Flexibility within a near-minimum set of quarter-degree grid cells selected to represent a group of 321 species of Madagascan butterflies at least once (ignoring other constraints). (a) Gray spots are the selected areas from Fig. 3. For selected area 13 in the northeast, the gray squares show richness in other cells of the four goal-essential species (Fig. 7) that justify its inclusion in the area set. Therefore, the other black square that shares all four of these goal-essential species is a fully flexible alternative (and could replace area 13), whereas the other gray squares are only partially flexible (more than one of these would be required to replace area 13). (b) The same map as in a, but showing the fully flexible areas for all selected areas. The black areas are either irreplaceable (Fig. 3) or have no fully flexible alternatives. The other gray areas are now groups of fully flexible alternatives, so only one area is required from each group. For example, the two black squares sharing all four goal-essential species from a are now shown in dark gray, as a group of fully flexible areas for selected area 13. From the numbers of fully flexible areas in each case, the total number of alternative near-minimum sets is estimated as $1^{23} \times 2 \times 2 \times 2 \times 10 \times 15 \times 32 = 38{,}400$ (see Williams *et al.*, 2000) (data courtesy of David Lees).

representative area sets, whatever the size or cost of the set. This depends on the rarest species in each area because the rarer a species, the fewer the alternative areas available to represent it. For this purpose, "rarity" of a species is measured from its frequency of occurrence among the areas that are candidates for selection. An approximate estimate of shades of irreplaceability is "effective maximum rarity" (EMR), which depends on the rarest species in each area, but considering only the species that are complementary to areas with even rarer species (Pressey *et al.*, 1994):

$$\mathrm{EMR} = 100/c_{ij}$$

where c_{ij} is the frequency of the rarest unrepresented species (i) that is present in an area (j). If no species are left unrepresented, then

$$\mathrm{EMR} = 100/C$$

where C is the total number of candidate areas in the data. This measure of irreplaceability captures much of the relational property of complementarity among areas in a single score.

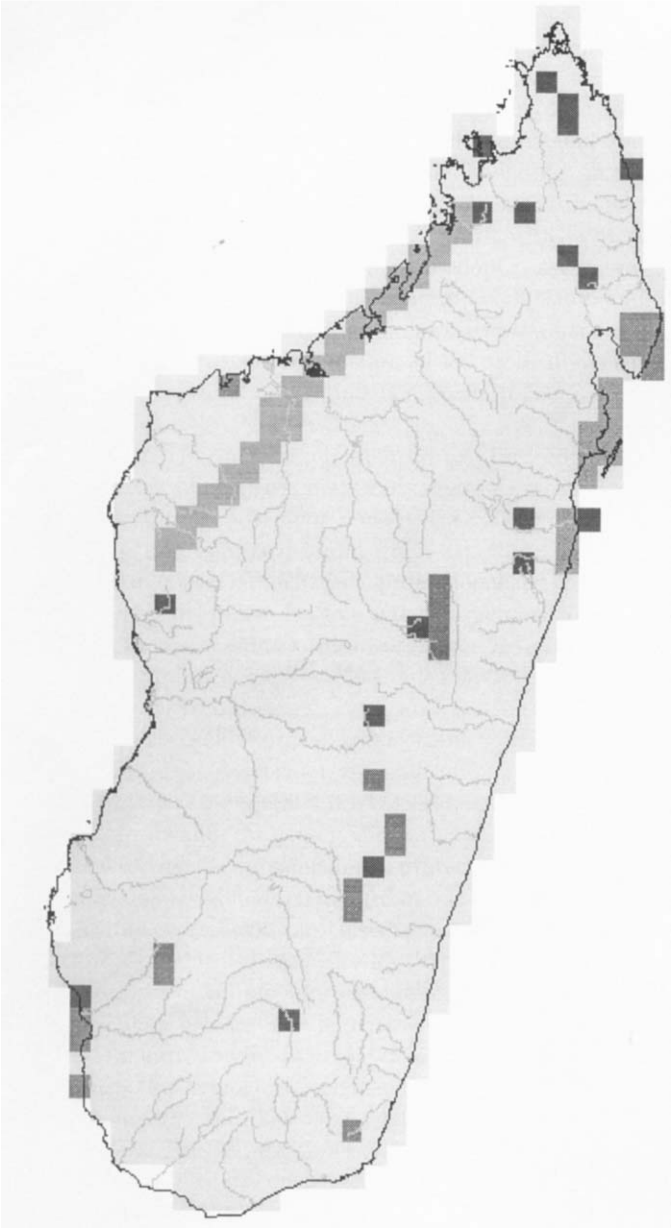


FIGURE 9 "Shades of irreplaceability" estimated using effective maximum rarity scores for a group of 321 species of Madagascan butterflies among quarter-degree grid cells. Black cells show maximum scores, dark gray lower scores, and light gray minimum scores (data courtesy of David Lees).

Using the definitions at the beginning of this article, shades of irreplaceability is a measure of flexibility (Fig. 9) rather than a measure of the conservation value of an area (which would be measured in terms of an appropriate part of its biodiversity value). When the EMR score is applied in a stepwise manner to select areas, it is then almost identical to the progressive rarity heuristics described previously (Fig. 5). Indeed, the EMR score may suffer from similarly reduced efficiency if it is not also combined with appropriate redundancy checks, although more efficient procedures for estimating shades of irreplaceability are being sought. Once this is achieved, shades of irreplaceability will be very similar to flexibility but expressed in relation to all possible fully representative sets of areas rather than just minimum-cost sets of areas. Whichever term is used, this approach can be used as another way of prioritizing area choices against threat by their number of flexible options (see Section III,D).

D. Comparing Area Sets

Two kinds of assessments of area sets require special care. The first is the comparison of sets of complementary areas with other sets of areas. The second is counting the representation of species or area attributes other than those used to select the areas. The complication arises from flexibility among alternative complementary area sets, which may result in different degrees of overlap among area sets or in different degrees of representation of the other organisms. In both cases, the flexible areas cannot be ignored because they are being compared on the basis of attributes other than those for which they were selected, and which therefore may be distributed among areas quite differently. For example, in Fig. 10, the flexible sets of 50 areas selected for maximum coverage of all mammals and birds must by definition all have the same number of mammals and birds. However, when sets of 50 areas are selected for maximum coverage of just the flagship species, or just the large mammals, then any flexible sets may vary in the numbers of all mammals and birds represented.

When assessing representation of these other area attributes for a set of complementary areas that shows flexibility, an area set has to be treated as a population of area sets. This gives rise to a corresponding distribution of differing degrees of area coincidence or of species representation. If two complementary area sets are being compared and both show flexibility, then the full intersection of the two populations of sets has to be considered.

V. CONCLUSIONS

Complementarity allows the maximum amount of biodiversity value to be represented for conservation, sub-

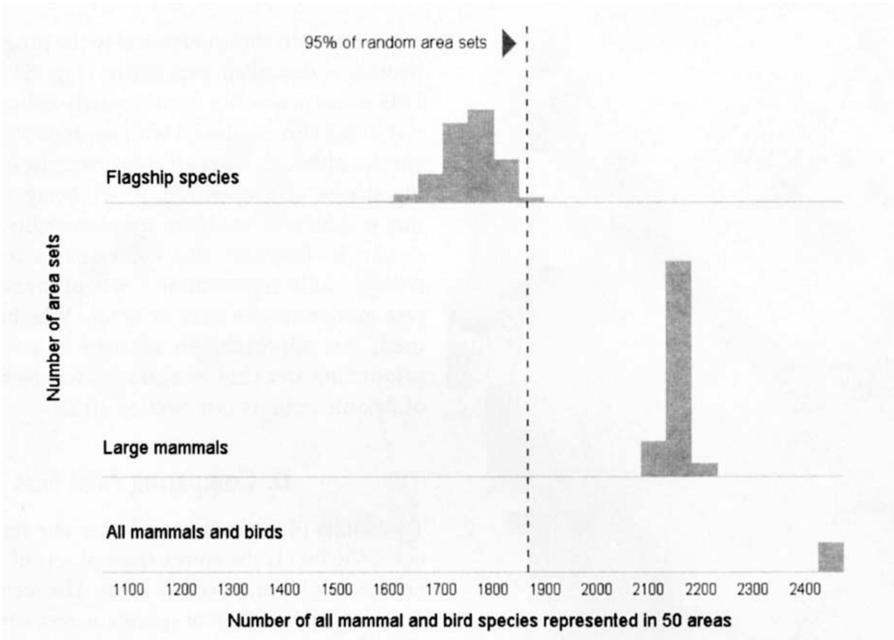


FIGURE 10 Flexibility and representation of mammals and breeding bird species. Histograms of representation of all mammals and breeding birds within flexible sets of 50 one-degree grid cells (each approximately 11,025 km^2) in sub-Saharan Africa, each selected for maximum coverage of three different groups of species: upper, six flagship mammals; middle, five orders of larger mammals; and lower, all mammals and breeding birds. With the exception of the histogram at the bottom, which shows all flexible sets of 50 cells, scores are from a randomly drawn sample of 1000 alternative sets from an estimated 2.12×10^{86} flexible sets of 50 cells for flagship species and an estimated 6.54×10^{41} flexible sets of 50 cells for larger mammals. Representation scores to the left of the dashed line are within the range expected when choosing 50 grid cells at random (adapted from Williams *et al.*, 2000).

ject to the constraints specified in the goals of an analysis, which inevitably include limited resources. Several different techniques are commonly used to achieve this end. However, the particular techniques are not the essential feature, and the approach can be much more flexible for dealing with diverse and heterogeneous goals and data than simple test exercises may be taken to imply. What is important is that, in principle, complementarity can be exploited in combination with any available and appropriate information on the distribution of diversity, on its viability, on threats, on costs, and on any other ecological or social constraints that are currently being used to make decisions. Consequently, it should be applicable to any situation in which we wish to conserve as much biodiversity as possible in order to keep track of complex data for improving on ad hoc decisions. The major challenges for the immediate future are how best to predict viability and threat from limited information and how best to integrate this information into complementarity-based area-selection methods.

See Also the Following Articles

CONSERVATION EFFORTS, CONTEMPORARY • SPECIES-AREA RELATIONSHIP

Bibliography

Bedward, M., Pressey, R. L., and Keith, D. A. (1992). A new approach for selecting fully representative reserve networks: Addressing efficiency, reserve design and land suitability with an iterative analysis. *Biol. Conserv.* 62, 115–125.

Camm, J. D., Polasky, S., Solow, A., and Csuti, B. (1996). A note on optimal algorithms for reserve site selection. *Biol. Conserv.* 78, 353–355.

Csuti, B., Polasky, S., Williams, P. H., Pressey, R. L., Camm, J. D., Kershaw, M., Kiester, A. R., Downs, B., Hamilton, R., Huso, M., and Sahr, K. (1997). A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biol. Conserv.* 80, 83–97.

Faith, D. F., and Walker, P. A. (1996). Integrating conservation and development: effective trade-offs between biodiversity and cost in the selection of protected areas. *Biodiv. Conserv.* 5, 431–446.

Kirkpatrick, J. B. (1983). An iterative method for establishing priorities for the selection of nature reserves: An example from Tasmania. *Biol. Conserv.* 25, 127–134.

Margules, C. R., Nicholls, A. O., and Pressey, R. L. (1988). Selecting networks of reserves to maximise biological diversity. *Biol. Conserv.* 43, 63–76.

Pressey, R. L., and Nicholls, A. O. (1989). Efficiency in conservation evaluation: Scoring versus iterative approaches. *Biol. Conserv.* 50, 199–218.

Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I., and Williams, P. H. (1993). Beyond opportunism: Key principles for systematic reserve selection. *Trends Ecol. Evol.* 8, 124–128.

Pressey, R. L., Johnson, I. R., and Wilson, P. D. (1994). Shades of irreplaceability: Towards a measure of the contribution of sites to a reservation goal. *Biodiv. Conserv.* 3, 242–262.

Pressey, R. L., Possingham, H. P., and Margules, C. R. (1996). Optimality in reserve selection algorithms: When does it matter and how much? *Biol. Conserv.* 76, 259–267.

Pressey, R. L., Possingham, H. P., Logan, V. S., Day, J. R., and Williams, P. H. (1999). Effects of data characteristics on the results of reserve selection algorithms. *J. Biogeogr.* 26, 179–191.

Rebelo, A. G. (1994). Using the Proteaceae to design a nature reserve network and determine conservation priorities for the Cape Floristic Region. In *Systematics and Conservation Evaluation* (P. L. Forey, C. J. Humphries, and R. I. Vane-Wright, Eds.), pp. 375–396. Oxford Univ. Press, Oxford.

Vane-Wright, R. I., Humphries, C. J., and Williams, P. H. (1991). What to protect? Systematics and the agony of choice. *Biol. Conserv.* 55, 235–254.

Williams, P. H. (1998). Key sites for conservation: Area-selection methods for biodiversity. In *Conservation in a Changing World* (G. M. Mace, A. Balmford, and J. R. Ginsberg, Eds.), pp. 211–249. Cambridge Univ. Press, Cambridge, UK.

Williams, P. H., Burgess, N. D., and Rahbek, C. (2000). Flagship species, ecological complementarity, and conserving the diversity of mammals and birds in sub-Saharan Africa. *Anim. Conserv.* (in press).

COMPLEXITY VERSUS DIVERSITY

Shahid Naeem
University of Washington

GLOSSARY

biotic interaction When the growth or behavior of one species affects those of another species; such interactions can be antagonistic (e.g., competition for limiting resources, predation, herbivory, or parasitism) or facilitory (e.g., pollination or other forms of mutualism).

community Set of coexisting species in an ecosystem.

food web Set of trophic relationships among species in a community; schematically, the pattern often resembles a web of species each connected by trophic interactions with other species.

functional dependency Dependency of one species on another to complete a particular ecosystem process or function; an example would be the dependency of plants on decomposers and decomposers on plants to recycle nutrients.

phylogeny Pattern of evolutionary descent among species, generally illustrated as a "tree" in which nodes represent common ancestors and lengths of branches represent time since divergence; time since divergence is generally assumed to correlate with genetic relatedness.

process intricacy Complexity of temporal or spatial patterns of ecological processes such as population dynamics or production.

structural intricacy Complexity of patterns of links or connections among species in a community that are created by biotic interactions, shared pathways of nutrient and energy flow, or phylogenetic relationships.

trophic interaction Feeding relationship between two species; these include predation, herbivory, parasitism, bacterivory, frugivory, or any other interaction that involves individuals of one species consuming individuals or parts of individuals from another species; trophic interactions represent a subset of all biotic interactions in a community.

ECOLOGICAL COMPLEXITY AND BIOLOGICAL DIVERSITY are often presumed to be strongly correlated or even synonymous with one another, but there is no necessary relationship between these terms. Ecological complexity is a scale by which we compare ecosystems. At one end of the scale, systems that exhibit unpredictable behavior or many connections among its species are considered "complex." At the other end of the scale, systems that exhibit highly predictable behavior or largely independent species are considered "simple." In

Encyclopedia of Biodiversity, Volume 1

contrast, biological diversity is a measure of the extent of genetic and phenotypic variation found within a community. These two characteristics of ecosystems, complexity and diversity, though fundamentally related, are clearly distinct. Although diversity and complexity are not necessarily correlated, the potential for a community to exhibit complexity is likely to be associated with its diversity. Widespread changes in diversity experienced by most ecosystems have generated considerable interest in quantifying complexity and assessing whether current changes are affecting levels of ecological complexity in natural ecosystems. The resolution of issues surrounding the relationship between complexity and diversity is an active area in contemporary ecological research.

I. THE COMPLEXITY OF ECOLOGICALLY DIVERSE SYSTEMS

> Hence, if certain insectivorous birds (whose numbers are probably regulated by hawks or beasts of prey) were to increase in Paraguay, the flies would decrease—then cattle and horses would become feral, and this would certainly greatly alter (as indeed I have observed in parts of South America) the vegetation: this again would largely affect the insects; and this, as we just have seen in Staffordshire, the insectivorous birds, and so onwards in ever increasing circles of complexity.
>
> —CHARLES DARWIN, 1859

> It is reasonable to believe that in following the numbers of orders in the fossil record we are indeed following the approximate overall course of ecological complication and diversification even though in a necessarily loose way.
>
> —G. G. SIMPSON, 1969

> Roughly speaking, we here take complexity to be measured by the number and nature of the individual links in the food web.
>
> —R. M. MAY, 1974

> The complexity of a system is in the eye of the beholder. It is measured by how well we understand causes, expect behaviours and, in praxis, achieve purposes. Hence, large numbers of variables, non-linear relationships among them, and the open nature of a system are important only to the degree they present barriers to understanding.
>
> —C. S. HOLLING, 1985

A. The Complexity of Nature and the Entangled Bank

Darwin provided us with a wonderful metaphor for what is intuitively meant by "ecological complexity" in his famous closing passage to *The Origin of Species*. In this passage, Darwin envisioned a riverbank teeming with a diverse and entangled array of organisms all involved in a complex web of biotic interactions. At the heart of this seemingly complex system, however, lies evolutionary and ecological processes that produce and maintain such remarkable diversity. For Darwin, biodiversity and the complexity that accompanies it are what distinguish nature from the inanimate features and processes of our world.

The entangled bank metaphor highlights two main ingredients of ecological complexity. First, there is biodiversity itself, with a clear implication that greater diversity provides a greater potential for complexity. Second, and more importantly, there is the web of interactions among species, with an equally clear implication that more complex webs (e.g., greater numbers and kinds of biotic interactions among species) are indicative of more diverse communities.

This metaphor also highlights two distinct meanings of complexity. First, complexity refers to the intricacy of the pattern of connections among species. This is the meaning used in many studies that have related ecological complexity to ecological stability. The second meaning applied to the term complexity is when nature defies easy characterization and cannot be reduced to a finite set of processes and principles that govern its order. This second meaning is used by the "macroecological" perspective (Brown, 1995), the hierarchical perspective of nature (O'Neill *et al.*, 1986), when considering nature as complex adaptive systems (Levin *et al.*, 1997), or when nature is viewed as a self-organizing system (Holling, 1992).

Figure 1 illustrates these different kinds of complexities. The top row of Fig. 1 shows the dynamics of an ecosystem process as an example of an ecosystem property that may exhibit simple or complex behavior. This process could be production, decomposition, or any other process representing the aggregate activities of species in the community. The remainder of Fig. 1 shows three fundamental kinds of linkages among species in communities: community structure or the biotic interactions among species, phylogenetic structure or the pattern of phylogenetic relationships among species, and functional structure or the pattern of functional dependencies among species. The pattern of tem-

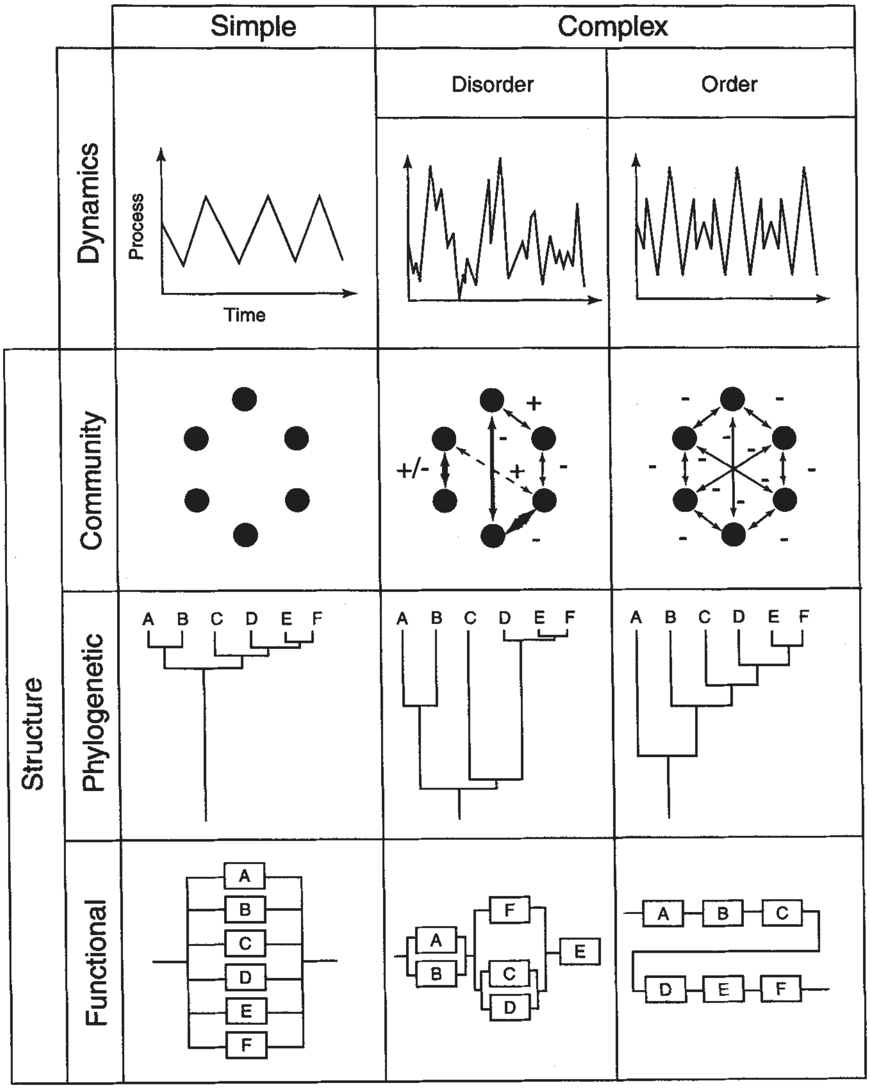


FIGURE 1 Simplicity and complexity in ecological systems. Dynamics are illustrated as temporal patterns in an ecological or evolutionary process, such as production, standing biomass, energy or nutrient flow, or rates of diversification. Community structure is illustrated as the number, type (e.g., competition or facilitation, shown as − or +, respectively), strengths (thickness of arrows), and arrangement of interactions (arrows) among species (circles). Phylogenetic structure is illustrated as phylogenetic trees in which every two species share a common ancestor. The vertical lengths of branches indicate distances, assumed to be a measure of genetic similarity or time since divergence. The complex phylogenies of more distantly related species contain more genetic information than would be found in the simple phylogeny of many closely related species. Functional complexity is illustrated as species whose contributions to ecosystem processes are either largely dependent (serially linked) on the activities of other species or exhibit a diverse array of interdependencies among one another. A simple system is one in which all species contribute to the functioning of an ecosystem in the same way (parallel redundancy) such that the loss of a species from a system does not affect its functioning.

poral or spatial variation of processes is known as process intricacy. The pattern of linkages, whether community, phylogenetic, or functional, is known as structural intricacy.

Whether considering process or structural intricacy, exhibiting either highly ordered states or highly irregular or disordered states is considered complex. As shown in Fig. 1, a temporal pattern for a specific ecological process that exhibits highly ordered, predictable, repeated series of oscillations would be considered complex. Yet a highly irregular, unpredictable, chaotic pattern of oscillation would also be considered complex when compared to a simple, mildly oscillating process. Similarly, Fig. 1 shows that a pattern of biotic interactions in which every species is linked to every species in the community by exactly the same kind of interaction with exactly the same strengths of interactions creates a symmetrical, highly ordered structure that would be considered complex. At the same time, a community that has a jumbled network of biotic interactions that vary in type and strength would also be considered complex when compared to a simple set of independent, noninteracting species.

Although ecological complexity is difficult to define precisely given the many different ways the term is used, when we consider the basic elements of ecological complexity it becomes apparent that diversity is not a necessary part of the definition of complexity. Note in Fig. 1, for example, that all nine structures shown are made up of six species. In spite of having the same number of species, these structures differ quite dramatically in complexity.

Despite the fact that diversity is not a necessary part of the definition of complexity, diversity and complexity are related in an elementary yet fundamental way. Few would debate that extremely depauperate ecosystems such as monoculture plantations of bananas are less complex than the extremely species-rich tropical rain forests they replaced, but for the bulk of ecological communities that lie somewhere between these extremes of biodiversity, determining just how complex a given system is can be quite difficult. For example, if a community contained 25 species of grasses while another contained only 10 species, but was made up of a mixture of 3 grasses, 4 forbs, and 3 shrubs, which community would be more complex? If a community consisted of 25 noninteracting species and another contained 5 species involved in a complex web of interactions, which would be more complex? If two communities each contained 10 species, but one exhibited chaotic dynamics while the other hovered close to a constant number of individuals per year in spite of repeated perturbations, which would be more complex?

Another way to look at the problem is to plot, simply for illustrative purposes, complexity against biodiversity as done in Fig. 2. Aside from the trivial points where there is no diversity, and therefore no ecological complexity, and where there is some diversity and there is some complexity, a variety of relationships could be drawn between the two points depending on one's perspective or experience. That is, complexity and diversity could either increase hand-in-hand or show any possible relationship that is bounded by a positive, asymptotic relationship between diversity and complexity on one end and an exponential relationship between diversity and complexity on the other end (see Fig. 2). Of interest would be whether some general relationship could be described or whether every ecosystem would show a unique relationship. If the latter is true, then knowing a system's diversity would provide no means for predicting its complexity and whatever ecological behaviors might be a function of complexity.

Because ecological complexity is believed to be an important part of biologically diverse ecosystems and because diversity is currently rapidly declining, the role of ecological complexity is of considerable concern to ecologists. Does biodiversity loss mean loss of ecological complexity? If so, what are the ecological consequences of losing ecological complexity? This article reviews ecological complexity, its definition, its measurement, its relationship to biodiversity, and what is known about its role in ecosystem functioning.

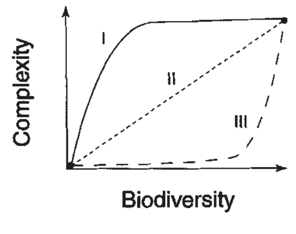


FIGURE 2 The potential relationships between complexity and biodiversity. Three possible relationships are shown. (I) The upper bound represented by an asymptote in complexity as biodiversity increases. (II) A linear relationship in which biodiversity is a direct measure of complexity in an ecosystem. (III) The lower bound represented by an exponential increase in complexity as diversity increases. Note that the general tendency is to quantify complexity as the number of interactions among species, and biodiversity as either species richness or the species richness weighted by relative abundance. Curve II represents the common perception of biodiversity and complexity being intimately related, but I and III represent the actual bounds for where the true relationship for a given ecosystem lies.

B. Concepts and Definitions

Concern over the potential alterations in ecological complexity caused by anthropogenic global change led the United Nations Environmental Program's Global Change and Terrestrial Ecosystem (GCTE) Core Project to establish Focus Four in 1996 to establish a research agenda for examining the ecosystem consequences of changes in ecological complexity. This program uses the following definition of ecological complexity:

> Ecological complexity represents biological diversity but in a broad sense, including not only species diversity but also diversity of ecosystems and landscapes, as well as genetic diversity within species. In addition, ecological complexity involves the diversity of trophic pathways and interactions, that is, the connectivity of ecosystems. We can envision ecosystems with similar diversity but contrasting complexity as a result of different organizational structures.
>
> —SALA, 1996

The first part of this definition of biodiversity was put forth in Article 2 of the 1992 Convention on Biological Diversity. In this Convention, biodiversity was defined as "the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic systems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems" (Harper and Hawksworth, 1995). This virtual synonymy with the first part of the GCTE definition of ecological complexity clearly indicates how strong a sense there is that biodiversity and ecological complexity are tightly correlated.

The second part of the GCTE definition concerns biotic interactions, which equates linkages among species with complexity. "Complexity," however, is not clearly defined, nor is it apparent from such a definition how one might go about measuring ecological complexity. Complexity is actually quite difficult to define and as a concept it has received far less attention than other subjects in the philosophy of science, but if we are concerned about its significance in ecological systems it would be valuable to begin by clearly delineating just what exactly is meant by "ecological complexity." In most instances, it is largely a matter of judgment that one thing is considered more complex than another, a judgment often based on a fair number of factors rather than just one or two. Thus it is not surprising, as discussed earlier, that species richness alone will fail to capture what is meant by ecological complexity.

Chapman (1985), in his review of a symposium on the concept of complexity, identified four distinct meanings of "complex" used in the conference and in the literature. First, there is what Chapman called "complication," or the process by which a system becomes naturally complex, but such complexity is ultimately "reducible" to fundamental principles or mechanisms. Second, there is "complification," which, like complication, concerns reducible systems, but involves purposeful design (e.g., a watch or a banana plantation has a different kind of complexity than the solar system or a tropical rain forest). Third, there is "complexification," which concerns irreducible systems (systems that are not composites of simpler parts). Finally, the fourth usage is "complexity" itself, which is simply a state of uncertainty concerning whether a system is reducible or not (irrespective of it being natural or of purposeful design). These decidedly ugly terms are not likely to catch on, but they do point to the variety of ways that ecologists, like other scientists, use complexity.

C. Community, Phylogenetic, and Functional Complexity

Typically, the web of interactions within a community, characterized by the number, nature, strengths, and structural pattern of biotic interactions among species, is often what is treated as ecological complexity. This can be the web of all interactions (community web, including competition, facilitation, parasitism, predation, and other interactions) or a subset of the community web, such as just trophic or feeding interactions (food web, including predation, herbivory, disease, or other consumer–resource species interactions).

In contrast to community complexity, however, phylogenetic complexity may be of interest when a community contains a diverse array of taxa. From the perspective of systematists or paleontologists, simply having many species that are phylogenetically diverse is evidence for complexity (Raven, 1996). A system of 15 species of grasses from the same genus might be considered less phylogenetically complex than a system of 10 species from 10 genera of grasses, which in turn might be less complex than a system with 5 species each from a different plant family.

Finally, complexity of an ecological system may be tied to its functioning as a biogeochemical system rather than to its population or biomass dynamics. That is, the flow of nutrients and energy through the system may be of interest rather than the dynamics of the many populations found within it. Note that ecosystems function in the sense that they exhibit activity or processes such as nutrient and energy flow. "Function" need not imply purpose or design (an unfortunate connotation of the term "functioning").

From the foregoing, it is apparent that ecological complexity is a composite feature of ecosystems consisting of an extraordinary number of different factors, including population, phylogenetic, and functional characteristics of the species in the community. The full measure of the complexity of a single ecosystem would therefore include a large number of factors, most of which are difficult to quantify. In general, increases in any of the properties listed in Table I, whether associated with structural or process intricacy, tend to increase ecological complexity, whereas decreases tend to simplify them. Table I demonstrates that measures of biotic diversity per se, such as species richness or evenness, are likely to be inadequate measures of complexity, though they may serve as crude proxy measures. The next section considers some basic measures of complexity, but not surprisingly, it will be evident that such measures of complexity generally deal with only single dimensions of complexity.

TABLE I

Measures of Ecological Complexity

Term	Measure
Structural intricacy	Numbers of populations
	Numbers of species
	Numbers of biotic interactions
	Kinds of interactions
	Strengths of interaction
	Higher-order effects
	Indirect effects
	Connectedness
	Number of patches
	Arrangement of patches
	Immigration probabilities
	Emigration probabilities
	Number of functional groups
	Number of functionally singular species (inverse of redundancy)
	Number of links among functional groups
	Number of compound functional groups (e.g., omnivores)
	Phylogenetic distances among species
	Phylogenetic moment
Process intricacy	Stability
	Predictability
	Reliability
	Degree of nonlinearity
	Succession
	Population growth
	Ecosystem processes
	Extinction
	Speciation

II. MEASURES AND CONSEQUENCES OF ECOLOGICAL COMPLEXITY

A. Measures of Complexity

Because complexity is a scale by which we compare communities and ecosystems, quantifying or measuring complexity is important in providing accurate means of making comparisons. Measuring complexity, however, is difficult because complexity is difficult to define. Boulding (1985) nicely described the problem at a United Nations conference held in Tokyo on the theory of complexity. He stated:

> Complexity is a very complex concept, which is not altogether surprising. We do have a vague concept of it as a somewhat linear property, both of structures and processes, in which we make judgments of "more" or "less." I am pretty sure I am more complex than an amoeba.... Exact measurement of the concept eludes us and probably always will, for although we do make these estimations of it in terms of more or less, it represents a highly multidimensional reality that cannot be reduced to a linear measure.

In spite of the obvious difficulties Boulding outlined, ecologists and evolutionary biologists have nevertheless provided a number of measures of complexity, which are reviewed in the following sections. These measures can be used in conjunction with measures of diversity to assess the relationship between complexity and diversity.

B. Measures of Community Complexity

Measures of ecological complexity are closely allied with measures of diversity, which reinforces the common practice among ecologists of treating biodiversity and complexity as one and the same. At its simplest, one might consider diversity to simply be the number of species (S), or species richness. Such a measure is often felt to be inadequate because it does not reflect differences in commonness and rarity nor the degree of connectedness among species. Thus, many other measures of diversity are used that attempt to take linkages or potential linkages into account. For example, the Shannon-Weaver index (*H*) considers both species richness and the relative abundances of species (evenness) as the important components of diversity. *H* quantifies diversity by assuming that richness and evenness directly represent the potential number of community configurations that a particular assemblage of species can produce. The original index quantified information, summing logged probabilities of a system having particular configurations.

The configuration of a community is seldom known, let alone the variety of possible configurations it can occupy or the probabilities of taking on these configurations. One therefore uses proportional abundances (*p*) of species as a means of approximating probabilities of configurations, a strategy that some consider fairly unsound.

The formula using this approach is

$$H = -\sum_{i=1}^{S} p_i \ln p_i \tag{1}$$

where *i* refers to the *i*th species in the community.

A related measure of potential complexity that compares the full potential with the realized potential for configurations is called connectivity (*C*), or the proportion of actual interactions (links or connections) to total possible interactions, which is distinct from food web connectance (see later). The formula for connectivity is

$$C = \sum_{i \neq j} a_{ij} p_i p_j \Big/ \sum_{i \neq j} p_i p_j \tag{2}$$

in which *p* is the probability of a species being present, and a_{ij} is the probability of interaction when two individuals, each from a different species (*i* and *j*), encounter one another. In reality, identifying interactions in biological communities is very difficult and cannot be deduced from abundance data. Indeed, Margalef and Gutiérrez (1983) explored connectance by examining components and "interactions" in functional plans for construction toys because information of equivalent clarity from real communities was unavailable.

The Shannon formalism serves as the basis for the preceding and other measures of complexity and diversity, but this practice is questionable. The Shannon formalism quantifies complexity by estimating an upper bound for the number of configurations possible for a given set of entities as if all configurations would be functional. In reality, however, rules of assembly ultimately determine the actual limits of information that a set of entities can store and retrieve efficiently. For example, if one had a box of 26 unique letters (i.e., A–Z), each equally abundant, and a box of 4 unique nucleotides, each equally abundant, the Shannon formalism would consider the box of letters to have a greater capacity to store and retrieve information. Assembly rules for letters, however, must be followed to produce sentences containing useful information just as assembly rules for nucleotide sequences must be followed to produce useful proteins. It is quite possible that 4 nucleotides could code for a greater diversity of proteins in nature than 26 letters can code for words in English, the difference being attributable to the rules used in each system. Without knowing these rules, it is difficult to assess the actual information content for a collection of entities such as letters, nucleotides, or species. Although assembly rules for nucleotides and letters in human languages are well understood, assembly rules for species in communities, although well studied, are still poorly understood.

Appraisals of complexity based on static patterns of the distribution and abundance of species [Eqs. (1) and (2)] ignore flows of energy or nutrients and the dynamics of interacting populations, all of which are considered key elements of community complexity (see Fig. 1). Rutledge *et al.* (1976) adapted approaches based on information theory, as discussed earlier, to consider flows among species within a community. Their approach usefully distinguishes between and quantifies both flavors of complexity, that is, structured versus disordered patterns of flows among species.

Appraisals of community complexity may also be made using the pattern of biotic interactions that links species in a community. The notion that biotic interactions are the primary elements of structural intricacy (complexity) that determine community dynamics has been the dominant theme in community ecology for several decades, inspired by the original contributions of A. J. Lotka and V. Volterra. Mathematical treatments of complexity have explored the consequences of varying interactions by constructing model communities whose species richness, interaction coefficients, and signs and strengths of interaction vary randomly and by examining their stability features. The heart of this approach is to use the "community matrix" as a measure of complexity, a concept formalized by Richard Levins in 1968 and still very much in use today. The general approach is to assume that all populations can be modeled by differential equations of the form

$$\frac{dx_i}{dt} = r_i x_i (K_i - x_i - \Sigma\, \alpha_{ij} x_j)/K_i$$

where r_i is the intrinsic rate of increase of the population of the *i*th species, K_i is the carrying capacity of the *i*th species, x_i is the population size of the *i*th species, and α_{ij} is the effect of the *j*th species on the *i*th species. For a community at equilibrium, where

$$K_i = x_i + \Sigma\, \alpha_{ij} x_j$$

the system of equations can be written as

$$\mathbf{AX} \equiv \mathbf{K}$$

Here, **X** is the column vector of species population sizes, **K** is the column vector of species-specific carrying capacities, and **A** is the community matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdot & \cdot \\ \alpha_{21} & 1 & \alpha_{23} & \cdot & \cdot \\ \alpha_{31} & \alpha_{32} & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \qquad (3)$$

The properties of this matrix can then be used in analyses of complexity. Although wonderfully appealing as a logical measure of community complexity, identifying and quantifying interactions in a community is a formidable task and requires that a community be at or near equilibrium, rarely the case in nature.

One concern over measures of structural complexity based on biotic interactions is that they must account for "keystone" species. Defined originally by R. T. Paine in 1966 as a predator species that fosters the coexistence of competing prey species and thereby promotes structural complexity in ecosystems, it has become colloquially used to describe any species that has a disproportionate impact on the community in comparison to other resident species (Liebold, 1996) or species that serve as ecological engineers (Jones *et al.*, 1994).

Power *et al.* (1996) proposed that the distribution of "community importance values" (CIVs) may serve as a tool for comparing communities for this element of structural complexity. Formally, the *i*th species in a community has a community importance value (V) defined as

$$V_i = [(m_N - m_D)/m_N](1/p_i) \qquad (4)$$

where $m_N \equiv$ the quantitative property (e.g., production or rate of decomposition) of the intact community, $m_D \equiv$ the quantitative property of the community with the *i*th species removed, and p_i represents the proportional abundance of the *i*th species before it was removed. In practice, one would sequentially delete one species at a time, each time measuring the community response to such a loss. Eventually a distribution of CIVs would be obtained and statistics (e.g., the mean or modal CIV and its variance) could be used to compare one community with another. Communities with mean CIVs furthest away from 0 would be more complex than those whose distributions were not significantly different from a normal distribution with a mean of 0. Of course, such an experiment would be very difficult to conduct and Power *et al.* (1996) could identify only two such studies.

One approach to the intractability of contending with all species interactions is to focus on a subset of interactions. An example of such a focus is the study of feeding or trophic interactions. Typically, food web statistics serve to quantify food web structural complexity. These statistics include food chain length, connectance (the ratio of existing trophic links to possible links), compartmentation (the number of compartments or isolated subnetworks of interacting species in a food web), the ratio of predator species to prey species, and interval versus noninterval state (presence or absence of complex overlaps in resource sharing). The difficulty with all of these statistics is that they rest entirely on the quality of the set of food webs, many of which were not constructed for such analyses, and thus produce artifacts when subjected to such analyses.

C. Measures of Functional Complexity

From a functional perspective, complexity may be measured as simply the number of functional groups. Functional groups may be considered collections of species that are related by their ecosystem activities, though many other definitions have been used. For example, if a functional group were defined as the group of species sequestering carbon from the atmosphere by photosynthesis, then all macrophytes, algae, and cyanobacteria found in a lake would be considered part of this group, in spite of their very different ecologies and evolutionary histories. This degree of aggregation is much larger than what is typically found in contemporary food webs (trophic groups or species), community matrices (interaction coefficients from species pairs), or high-resolution phylogenies (orders, families, genera, species, etc.).

In the same way that tallying up numbers of species fails to capture complexity, tallying up functional groups is unlikely to serve as an adequate measure of functional complexity. More important would be the degree of parallel versus serial dependency among species in an ecosystem. As in community complexity, network architecture is the focus for measuring complexity, but rather than linkages of biotic interactions, only linkages of dependencies among species and functional groups become important.

Ecosystem reliability is an example of an index that combines measures of serial and parallel dependency (Naeem, 1998). Serial dependency occurs when every species or every functional group is dependent on the other for the ecosystem to function (e.g., decomposers depend on producers for energy while producers depend on decomposers for nutrient cycling). Parallel dependency occurs when one or more species within a functional group substitute for each other upon the extinction of one of the members of the group. In its simplest form, ecosystem reliability may be formally represented as

$$H(t) = \prod_{j=1}^{F} \left[1 - \prod_{i=1}^{S_j} (1 - e^{-\lambda t}) \right] \tag{5}$$

where F is the number of functional groups, S_j is the number of species in the jth functional group, and λ is the probability of extinction over a small increment in time, t. Note that the reliability of an ecosystem is defined as 1 [$H(0) \equiv 1$] and eventually decays in the absence of replacement of species by recolonization [$H(\infty) \equiv 0$].

As in other measures of complexity, the information needed for this index is difficult to obtain. Species, presence or absence, local extinction rates, and functional roles are often difficult to determine without considerable empirical work. Focusing on redundancy demonstrates, however, that if the properties being measured are the result of the collected activities of groups of similar (i.e., redundant) species, such as production, having many species could actually mean more steady levels of production (less complex behavior). Thus, the association between biodiversity and complexity of ecosystem functioning is dependent on the degree of species redundancy in the system.

D. Measures of Phylogenetic Complexity

Whereas community and functional complexity focus on structural aspects of ecosystems, phylogenetic complexity examines the structure of the phylogenies that relate species to one another within a community by their evolutionary history. As with community and functional complexity, the number of species is an important starting point, but the architecture of the evolutionary "tree" or "bush," as opposed to network architecture, serves as the primary basis for assessing phylogenetic complexity.

A straightforward measure of phylogenetic complexity might be simply the sum of the lengths in the tree that unites all species in the set being examined or a simple modification of the Shannon-Weaver index in which species are weighted by their taxonomic distinctness. More informative, however, is the "phylogenetic moment" [defined in Eq. (6)], which provides a statistic by which communities might be compared. A difficult necessity for using this measure is that one has to have the complete phylogenetic tree at hand (referred to as tree T), including those species that might have gone extinct and might be missing from the local community being examined. The current community will consist of some subset of T (referred to as set R), and the local community that one is measuring will invariably be missing some species and contain an even smaller set of species (referred to as set r). If we assume that the tree (its nodes and branches) can be represented by a series of points in time, with a total of PT points for the whole tree, then the phylogenetic moment for this subtree (M_R) is

$$M_R = \sum_{p \in P_T} d_{pr_p} \tag{6}$$

where d_{pr_t} is the distance between a point on the tree and the nearest species in the set r. The smaller the phylogenetic moment is, the more closely the community represents the full phylogenetic structure (found in T). For example, in Fig. 3, although the total distances among the species in trees R_1 and R_2 are the same, using the phylogenetic tree T, which includes extinct species F and G and missing but extant species B, shows that R_1 would have a smaller phylogenetic moment than R_2 because C would have more near neighbors than D. That is, A, C, and E will presumably contain genetic

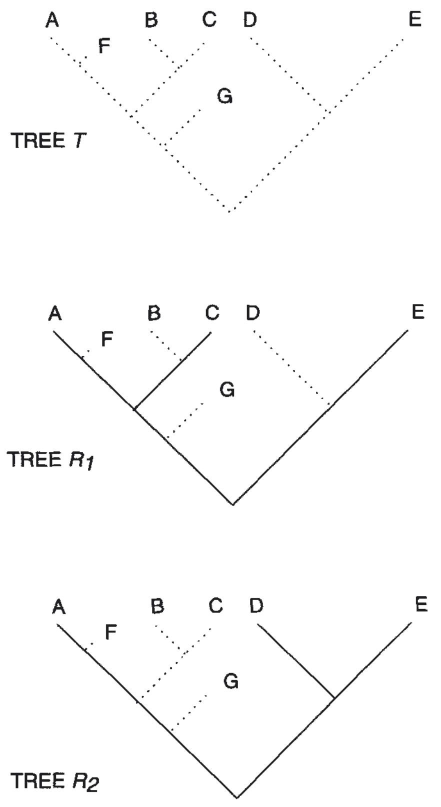


FIGURE 3 Hypothetical phylogenetic trees. Tree *T* represents the tree for all taxa in the community, including extinct species F and G. Two current local communities, each missing species B and one other extant species (dashed lines), are illustrated as R_1 and R_2. Solid lines show trees that would be used in calculating the total phylogenetic distance among species. Note that they are the same for R_1 and R_2. The phylogenetic moments for these trees would be the sum of the distances from each point on the tree *T* to the nearest species present. The moment would be smaller for R_1.

information more representative of the bulk of the species in the full tree (*T*) than *A*, *D*, and *E*.

E. Utility of Measures of Ecological Complexity

Although several measures of complexity have been developed and explored [Eqs. (1)–(6)], in reality no single measure is likely to adequately describe ecological complexity in all its forms. Virtually any index based on a subset of the factors listed in Table I is likely to produce rankings in complexity that will not agree with other methods, much in the same way that different diversity indexes may disagree in how they rank communities. The quantification of complexity is hindered by the difficulty in uncovering all species and all their interactions, their functional roles, and their phylogenetic relationships. As the next section shows, this means that we currently have only a partial picture of how complexity and biodiversity are related.

III. THE SIGNIFICANCE OF ECOLOGICAL COMPLEXITY

A. Complexity in Nature

Although diversity and complexity are practically universal features of ecosystems, we have much better information on diversity than on complexity. The estimation of existing diversity is, to be sure, challenging, but determining the distribution and abundance of species would only be the starting point for investigations of ecological complexity.

Unfortunately, the components of complexity are extraordinarily difficult to measure and for this reason empirical treatments have lagged behind the substantially greater progress made in theoretical explorations of the significance of ecological complexity. An ecologist never has a complete inventory of all species within a region, let alone their relative abundance, biotic interactions (direct, indirect, and higher order), phylogenetic relationships (at the regional level, including extinct species), functional roles, and population and process dynamics, or other details listed in Table I.

In most cases, however, the question at hand is more focused and thus more tractable. If the question concerns population stability, then quantifying only community complexity may suffice. If the question concerns what sets of species are most likely to contain the bulk of genetic information stored within a regional biota, then quantifying phylogenetic complexity may be enough. If the question concerns the magnitude or rate of ecosystem functioning, then quantifying functional complexity may do.

What is known about the relationship between diversity and complexity in nature? Several important patterns that emerge from studies that have examined community, functional, and phylogenetic complexity are considered here.

1. Community Complexity

An early and influential work that examined community complexity was R. H. MacArthur's 1955 study, inspired by E. Odum's theory that the "amount of choice which the energy has in following the paths up through the food web is a measure of the stability of the community." Using the Shannon-Weaver formula to model the number of alternative pathways and relate this to stability, MacArthur concluded that "stability increases as the number of links increases," though he was primarily considering resistance stability (remaining intact in spite of perturbation).

The theoretical relationship between complexity and stability is treated elsewhere in this volume, but it is worth noting here that this theory stimulated further investigations into the relationship between complexity and ecological processes, such as population dynamics. Later mathematical studies of local resilience stability (the ability to recover after perturbation) showed an inverse relationship between community complexity and resilience stability. Studies of connectivity and connectance (the nature and degree of connections among species) showed that resilience stability could either increase or decrease, depending on which trophic level was examined and the degree of connectance. Pimm (1979) also showed that complexity and stability could be positively related if one considered the deletion of species rather than resilience stability. Thus, those biologists who observed a positive association and those who observed a negative association between diversity and complexity more than likely differed in how they defined stability or how they defined complexity. Suffice it to say that there was agreement that complexity and stability were indeed related, but the nature of the relationship was dependent on one's definitions.

Although food web studies initially showed much promise as a means for examining the relationship between complexity (community structure) and diversity by focusing solely on trophic interactions, early findings did not fare well in the test of time. Polis and Winemiller (1996) summarize four patterns derived from early analyses of topographic or static food webs. First, there is constancy in the number of links per species in a food web irrespective of number of species. Second, the numbers of trophic levels seem to be invariably small (three or four). Third, the ratio of predator to prey species is relatively constant. Fourth, omnivory is rare. All of these patterns, however, have been challenged as artifacts.

Though potentially artifacts, food web patterns at least touched upon an obvious truth about communities—there are likely to be constraints in the structures of food webs. For example, at the simplest level, one cannot have consumers without resource species (i.e., no predators without prey) and there are likely to be stability or energy constraints that limit food chain lengths (Kaunzinger and Morin, 1998). Currently, food web biology has gone far beyond the sorts of patterns that initially stimulated research in this field, but new patterns, if they exist, await the completion of a new set of higher-resolution food webs.

The main observation to be drawn from studies of community complexity is that complexity is never at its maximum in diverse ecosystems. For example, although the potential number of pairwise interactions is large for any community, few of these interactions may ever be realized. As Margalef (1985) states, "the considerations of artifacts [from mathematical studies of diversity and complexity] provide an elementary but effective approach to the assertion that no system is completely interlocked or connected." We may infer from this fact that complexity asymptotes with diversity, an important consideration in our quest to understand the relationship between diversity and complexity (see Fig. 2).

2. Ecosystem Functioning

Biodiversity and ecosystem functioning are treated elsewhere in this Encyclopedia, but here I examine its relationship to complexity and functioning. May (1974) noted in his "afterthoughts" to the second edition of his influential book that aggregate properties of communities, such as production (i.e., ecosystem functions), could show the opposite relationship with complexity that dynamic stability showed. Theoretical support comes from different approaches, but experimental tests of the idea are few. Support for this idea can be found in Tilman and Downing's (1994) study of experimental grassland plots, though some controversy surrounds this study. In their study, production recovery from a drought (resilience stability) increased with plant diversity. Two laboratory microcosm experiments using freshwater microbial communities by J. McGrady-Steed *et al.* and S. Naeem and Li (both conducted in 1997) showed that predictability and reliability of ecosystem functions, such as production and community respiration, can indeed be associated with biological diversity and complexity.

3. Phylogenetic Complexity

The relationship between phylogenetic complexity and biotic diversity is not well understood. Globally, estimates of origination and extinction of species suggest more or less steady increases in biotic diversity punctuated by aperiodic mass extinctions. This steady rise may be taken to mean that origination rates, over millions of years, have been outpacing extinction, except during periods of mass extinction, and that recovery from mass extinctions has always occurred. Such a history of the biota suggests to some that the biota can tolerate even very deep "prunings" of the tree of life. At a global level, this suggests that random sets of species drawn from regional pools are likely to contain equivalent amounts of genetic information and that it takes severe levels of extinction to reach the point where insufficient genetic information is left in the remaining taxa to regenerate lost biodiversity. Because large-scale patterns in diversity across environmental gradients are similar across many taxa (Currie, 1991), we may expect that random samples from communities would result in phylogenies of similar complexity. Of course, since only a fraction of all species have been described, it is difficult to know exactly what the distribution and abundance of species are. As in community complexity and functional diversity, the true relationship between diversity and complexity is unknown, but it is likely to be asymptotic. That is, highly diverse systems may experience little change in phylogenetic complexity in the face of moderate levels of extinction.

V. CONCLUSIONS: THE DISENTANGLED BANK

Common usage of the term "complexity" in ecology implies that ecological complexity is a function of the phylogenetic, functional, and community structure of ecosystems. On the surface, ecological complexity and diversity seem so closely related that it is tempting to consider "biodiversity" as synonymous with "complexity." If we consider complexity to be the structural and process intricacies of a system, however, it is clear that diversity, in terms of numbers and types of organisms in a community, shows no necessary relationship with any measure of complexity aside from the trivial fact that without any diversity there is no complexity. Thus, two communities identical in numbers of species could be quite different in community, phylogenetic, or functional complexity. Likewise, a highly diverse system could be more or less complex than a depauperate system.

Studies of complexity in nature suggest asymptotic relationships between complexity and species richness, where species richness represents one way to define biotic diversity independent of complexity. It is hard to imagine, for example, that tropical rain forests, where species often number in the thousands, will show much reduction in complexity in the face of the loss of a few species. Current rates of extinction, however, are estimated to be staggeringly high and many diverse systems, such as tropical ecosystems, are not suffering extinction of species within their boundaries so much as they are being transformed to depauperate, managed ecosystems.

Thresholds in the response of ecosystems to the loss of diversity are particularly of interest (Levin, 1997) and these may be hypothesized from current knowledge about the relationship between diversity and complexity. From what is known about nature, both theoretically and empirically, it is reasonable to postulate that both community and phylogenetic complexity show asymptotic relationships with diversity. Because functional complexity aggregates many species into small numbers of functional groups (e.g., all plants may be considered producers), we can postulate that the asymptote for functional complexity will be lower than that for community or phylogenetic complexity. When these are plotted together, as shown in Fig. 4, we can identify two thresholds of interest. At some point, admittedly difficult to define precisely without accurate measures of community complexity or phylogenetic complexity, one will continue to experience no appreciable loss in functional complexity, but ecosystem properties, such as dynamic stability, may change and genetic resources will decline. As diversity declines even further, functionality is eventually lost after the second threshold (the leftmost dotted line in Fig. 4) is crossed.

Managed ecosystems, such as aquaculture systems, farms, or monoculture tree plantations, most likely reside at the lower threshold, where the managers maintain the system at its minimal diversity. In general, these ecosystems may be functionally complete, containing all necessary functional groups to ensure that the desired ecosystem functions are maintained or, if not functionally complete, humans provide missing functions, such as fertilizer addition where nitrogen fixers have

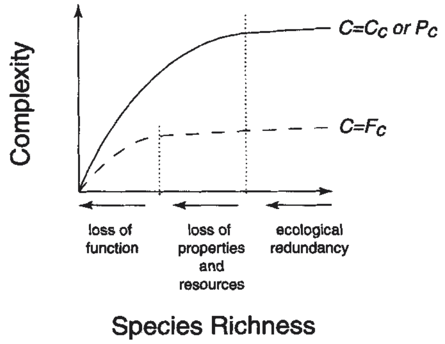


FIGURE 4 Hypothetical relationships between complexity and species richness. C ≡ complexity. Vertical dotted lines indicate where the threshold in ecological consequences of declining ecological complexity might be identified. C_C ≡ community complexity; F_C ≡ functional complexity; P_C ≡ phylogenetic complexity.

been removed. Complex systems are always more difficult to manage, but there are likely costs and consequences for resorting to simple systems, many of which have to do with system stability, system utility, and system efficiency, as the material in this article reveals.

The appropriate metaphor for the process of human transformations of ecological systems may be the disentangling of Darwin's entangled bank. As the earth's ecosystems are steadily transformed to agroecosystems and aquaculture systems, there is no question that average levels of biodiversity in ecosystems will continue to experience sharp declines in local biotic diversity, which eventually leads to losses in complexity. In many cases, understanding the ecological consequences of this loss may require focusing on ecological complexity, for which the quantification of species richness is only the first step.

See Also the Following Articles

DIVERSITY, COMMUNITY/REGIONAL LEVEL • FOOD WEBS • GENETIC DIVERSITY • LANDSCAPE DIVERSITY • MEASUREMENT AND ANALYSIS OF BIODIVERSITY • SPECIES DIVERSITY, OVERVIEW

Bibliography

Boulding, K. E. (1985). Learning by simplifying complexity: How to turn data into knowledge. In *The Science and Praxis of Complexity*, pp. 25–34. United Nations University, Tokyo.

Brown, J. H. (1995). *Macroecology*. University of Chicago Press, Chicago.

Chapman, G. P. (1985). The epistemology of complexity and some reflections on the symposium. In *The Science and Praxis of Complexity*, pp. 357–374. United Nations University, Tokyo.

Currie, D. J. (1991). Energy and large-scale patterns of animal– and plant–species richness. *Amer. Naturalist* **137**, 27–49.

Harper, J. L., and D. L. Hawksworth. (1995). Preface. In *Biodiversity Measurement and Estimation* (D. L. Hawksworth, ed.), pp. 5–23. Chapman and Hall, London.

Holling, C. S. (1992). Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecol. Monogr.* **62**, 447–502.

Jones, C. G., J. H. Lawton, and M. Shachak. (1994). Organisms as ecosystem engineers. *Oikos* **69**, 373–386.

Kaunzinger, C. M. K., and P. J. Morin. (1998). Productivity controls food-chain properties in microbial communities. *Nature* **395**, 495–497.

Levin, S. A. (1997). Biodiversity: Interfacing populations and ecosystems. In *Biodiversity: An Ecological Perspective* (T. Abe, S. A. Levin, and M. Higashi, eds.), pp. 277–288. Springer-Verlag, New York.

Levin, S. A., B. Grenfell, A. Hastings, and A. S. Perelson. (1997). Mathematical and computational challenges in population biology and ecosystem science. *Science* **275**, 334–343.

Liebold, M. A. (1996). A graphical model of keystone predators in food webs: Trophic regulation of abundance, incidence, and diversity patterns in communities. *Amer. Naturalist* **147**, 784–812.

Margalef, R. (1986). Ecosystems: diversity and connectivity as measurable components of their complication. In *The science and praxis of complexity* (U. N. University, ed.), pp. 228–244. United Nations University, Tokyo.

Margalef, R., and Gutiérrez E. (1983). How to introduce connectance in the frame of an expression for diversity. *Amer. Naturalist* **5**, 601–607.

May, R. M. (1974). *Stability and complexity in model ecosystems*. Princeton University Press,, Princeton.

McGrady-Steed, J., Harris, P. M., and Morin, P. J. (1997). Biodiversity regulates ecosystem predictability. *Nature* **390**, 162–165.

Naeem, S. (1998). Species redundancy and ecosystem reliability. *Conservation Biol.* **12**, 39–45.

Naeem, S., and Li, S. (1997). Biodiversity enhances ecosystem reliability. *Nature* **390**, 507–509.

O'Neill, R. V., D. L. De Angelis, J. B. Waide, and T. F. H. Allen. (1986). *A Hierarchical Concept of Ecosystems*. Princeton University Press, Princeton.

Pimm, S. L. (1979). Complexity and stability: Another look at MacArthur's original hypothesis. *Oikos* **33**, 351–357.

Polis, G. A., and K. O. Winemiller. (1996). Food webs: What do they tell us about the world? In *Food Webs* (G. A. Polis and K. O. Winemiller, eds.), pp. 1–24. Chapman and Hall, London.

Raven, P. H. (1996). Biological complexity. In *The Emergence of Complexity in Mathematics, Physics, Chemistry, and Biology* (B. Pullman, ed.), pp. 251–265. Pontificia Academia Scientiarvm, Vatican City.

Rutledge, R. W., Basore, B. L., and Mulholland, R. J. (1976). Ecological stability: An information theory viewpoint. *J. Theor. Biol.* **57**, 355–371.

Sala, O. E. (1996). Global change and ecological complexity. In *Global Change and Terrestrial Ecosystems* (B. Walker and W. Steffen, eds.), pp. 341–345. Cambridge University Press, Cambridge, United Kingdom.

Tilman, D., and Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature* **367**, 363–365.

COMPUTER SYSTEMS AND MODELS, USE OF

Louis J. Gross
The Institute for Environmental Modeling, University of Tennessee

GLOSSARY

aggregation Combining several potentially separate components of a system to simplify analysis.

dynamic model Mathematical description of a system that has components that vary in time.

multimodel Single integrated model that links together models taking different approaches.

parameter Constant in a model that must be estimated from data, or assumed to be of a particular value.

MODELS IN A VARIETY OF FORMS play a critical role in advancing our understanding of natural systems. Models abstract basic principles and derive the implications of such abstractions. This provides a method to analyze alternative hypotheses about natural system responses and the mechanisms that underlie these responses. This article presents an introduction to various modeling approaches and discusses how these have been applied to analyze species abundance and distribution.

I. WHAT IS A MODEL?

Just as storytellers can take their audience on trips to faraway places and provide a glimpse at life in different cultures, scientists tell their stories about the way the world works by making models. Such models never provide a complete view of how the world works, but do give us glimpses that help us to piece together interactions between different parts of the world and the processes that connect them. These models take many forms, some being mostly verbal, others mostly qualitative and graphical, some phrased in various mathematical forms, and still others set up as collections of rules within a computer program.

Models provide maps of varying levels of complexity to help us understand the topography of science. There are coarse road maps that provide merely the outline of major arteries for traffic, telling us nothing about buildings or other features of the landscape, but providing an overview of the linkages between key components of a system. More elaborate models show us the buildings and the infrastructure that links these buildings—the sewer and power lines. Even more complex models would indicate the humans in each building, their occupations, and the flow of money or capital

Encyclopedia of Biodiversity, Volume 1

goods between them. Similarly, models at many levels of complexity can be useful for addressing different questions in the study of biodiversity. A very coarse model might analyze the effect of land use change worldwide on total species richness. A more complex model could consider particular regions and analyze differentially the changes in land use within them and the associated changes in species numbers. A still more complex model could consider the local dynamics of individual species and from this elaborate the dynamics of species numbers.

II. THE PURPOSES OF MODELING

A. General Objectives of Models

Which type of model one constructs depends on the questions being asked and the availability of data to construct a reasonable model. Though there are many specific purposes for constructing models (Haefner, 1996), these may be grouped into a few general objectives: Description, Mechanism, Prediction, and Control. These objectives are not mutually exclusive, so that descriptive and mechanistic approaches may be used to aid prediction and control.

1. Description

Sometimes all that might be desired is a simple description of a collection of data. For example, an average provides a single value to summarize a list of numerical data. This may be sufficient for some purposes that do not require a description of how much variation is in the data. To assess variation, a dispersion measure such as the variance would be needed. These summary statistics are coarse, ignoring many of the details in the data. Yet they do allow us to easily comprehend major differences between different data sets. Extensive species lists within certain taxa from two locations may be compared by considering just the total numbers of species in the two locations and the number of species in common between them. Such a summary may be sufficient for a comparison of the two locations, while ignoring details such as the diversity within the taxa included. Descriptive approaches may be much more complex than simply providing averages and variances. Indeed the field of exploratory statistics deals with methods to analyze and summarize multidimensional data (Jambu, 1991). A typical example would be methods of time-series analysis in which the histories of species numbers might be compared between two locations or correlated with the histories of anthropogenic actions in the locations.

2. Mechanism

If the objective is to provide an understanding of how a particular system operates, then it is necessary to take account of the processes that govern the system. While all such mechanistic models are descriptive at some level, the point here is to deal with the basic physical, chemical, and biological processes operating in the system. This requires including those processes that operate at a spatial and temporal extent appropriate for the problem one is addressing, and ignoring others. Thus, analysis of how alternative global warming predictions would affect worldwide biodiversity might include the geographic variation in the temperature predictions at a spatial extent of hundreds of square kilometers, but would no doubt ignore the microclimate variation of every square meter. Even if it were possible to characterize the meter-by-meter temperature differences predicted by the alternative warming trends, the lack of available detail on the species present at this detailed spatial resolution limits the utility of including such detail. A discussion of such scaling issues is included elsewhere in these volumes.

3. Prediction

Predictive models are of two general types: those that attempt to project the behavior of the system based on certain explicit assumptions, and those that attempt to forecast the future behavior. The difference is between what might be true in the future if certain assumptions hold (projection) and what will be true in the future (forecasting) (Caswell, 1989). In many biological situations, the forecasting problem is not even attempted, as it would involved taking account of a wide variety of unpredictable abiotic phenomena (e.g., hurricanes and droughts). It is often feasible to construct a model to project the future dynamics of a system based on current observations and particular reasonable assumptions about the interactions in the system. The majority of population models (discussed elsewhere in these volumes) are of this form, in which abiotic influences are not included. These models can project the future behavior of the population based on the biotic forces of demographics, genetics, and social structure within the population. Uncertainties associated with unpredictable phenomena can be taken into account by attempting to project just the mean and variance of the variables of interest (e.g., population size).

4. Control

When a system has one or several components that are under human control, either completely or in part, then a model can be used to help determine how to apply such a control to meet certain objectives. Examples of controls are harvest quotas, limits on fertilization or pesticide application, flows from a dam, limiting importation of potentially harmful invasive species, and land use zoning regulations. Examples of objectives are maximizing biodiversity, minimizing population extinction probability, reducing the spread of nonnative species, and maintaining population size above some determined threshold (such as a minimum viable population size). Control models mostly focus on the dynamics of the system, with the simplest form of control being bang/bang, meaning on/off, such as allowing harvest in certain years and not allowing harvest in other years. A related objective is for control models to produce a relative comparison of alternatives in order to rank these alternatives according to some criteria (DeAngelis *et al.*, 1998). Still other control models are used to analyze the physiological responses of individual organisms to varying environments and the homeostasis that can arise through these responses.

III. TYPES OF MODELS

Models can be physical, such as animal models used in drug testing and airplane models used in wind tunnels. In biodiversity contexts, microcosms and mesocosms, which are limited biological systems built in a laboratory setting, play this role. These are meant to mimic the key biotic forces interacting within a natural system, but are constructed at a spatial extent that allows for easy observation and controlled experimentation. They cannot include all of the components of the real system, but do allow for projection of how the real system might respond under particular perturbations. Physical models clearly are limited, particularly to organisms that are mostly sessile or have very short distance movements.

Mathematical models come in a wide variety of forms. Some are simply graphical relations that show the qualitative relationship between certain components of a system, mainly to demonstrate the shape of response and whether one component increases or decreases with another. An example would be the increase and then decrease in species diversity along a gradient from low to high frequency of disturbance (Fig. 1). Here, there is no attempt to predict at exactly what disturbance frequency the exact peak in species diversity occurs. Rather, the objective is to illustrate the qualitative behavior of diversity, representing the "intermediate disturbance hypothesis," which posits that highest diversity occurs at intermediate disturbance frequencies.

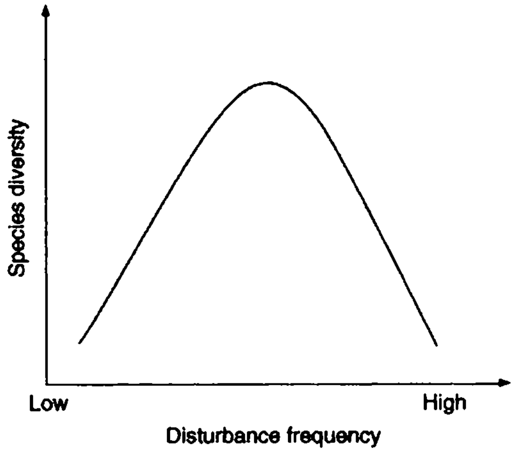


FIGURE 1 Illustration of the intermediate disturbance hypothesis.

The majority of mathematical models in ecology deal with the dynamics of populations and communities. Such models consider the basic processes of birth and death, immigration and emigration, and competition and predation to elucidate general theories of population dynamics. Described using differential or difference equations, these models allow for projection of the long-term behavior of populations, as well as provide methods to project the within-population structure (age, size, genetic, etc.).

Computer models are quite varied in structure. First, all the standard mathematical models of populations and communities, constructed using differential or difference equations, may be implemented on computers. Indeed, since it may be quite difficult to develop analytic solutions for such models, analysis of their behavior often requires the use of numerical solution methods implemented on a computer. There are many computer models that, although they may have a description that is essentially mathematical in form, are really described by the code itself rather than an explicit set of mathematical equations. An example would be cellular automata models, one type of which consists of a two-dimensional lattice, with each point on the lattice having one of a number of states (Langton, 1988). The simplest situation would be each lattice point being occupied (e.g., in the 1-state) or unoccupied (in the 0-state). The model is then described by a set of rules that determines how the state of a lattice point changes from one time step to another, based on the states of surrounding lattice points. Such a cellular automation

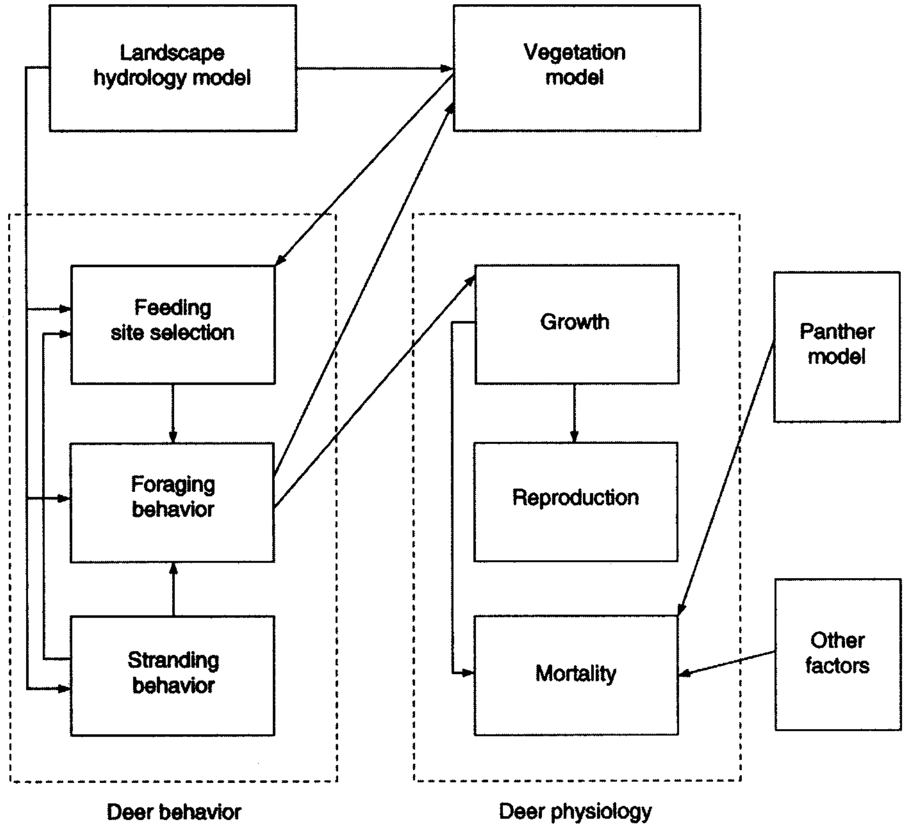


FIGURE 2 Graphical depiction of the major components of an individual-based model for deer. (Data from Comiskey *et al.*, 1997.)

may be used to mimic the spatial dynamics of populations, in which each lattice point represents a possible location of an individual. Alternatively, each lattice point can be interpreted as a local population, and the entire lattice then can follow the collection of such populations, called the metapopulation.

System simulation models are elaborate computer models that attempt to include most of the biotic and abiotic factors that affect the system. Many agricultural system models are of this type, and include the crop, its pests, soil nutrients, and weather conditions, among other factors. Some other types of computer models are described in later sections. In all cases, though the model is in essence specified by the code itself, it is very useful to have some graphical description of the major components of the model. One example is shown in Fig. 2. There are a number of general modeling software packages designed explicitly to aid construction of computer models through the use of graphical elements.

IV. LIMITATIONS OF MODELS

A. Trade-offs: Generality, Precision, Realism

No one model can do everything. In the process of deciding what components of a system to include, what processes to consider, and what spatial and temporal extent is appropriate, the model excludes part of reality. Modeling is a process of selective ignorance. We decide what to include and what to exclude. Part of the art of modeling is coming to grips with the issue of which details are important and which ones are not. In most cases the process is iterative, with a sequence of different models being tried until a model is arrived at that includes just the essential details necessary to address the problem of concern.

One view of the trade-offs in constructing a model is that no one model can be simultaneously general, precise, and realistic (Levins, 1968). As Fig. 3 illus-

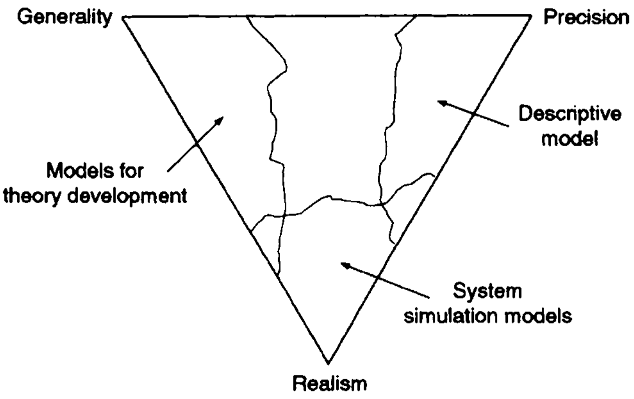


FIGURE 3 Trade-offs in modeling.

trates, these properties may be viewed as points of a triangle. Generality implies that the model may be useful in many different natural systems. A realistic model is one with components, parameters, and variables that are all possible to estimate from observations. A precise model is one that produces quantitative, accurate descriptions of the natural system. Models for theory development, including most of the classical population and community models, are quite general, somewhat realistic, but lacking in precision. Descriptive models designed to mimic the response of particular systems tend to be quite precise, slightly realistic, and not at all general. Most regression models are of this type. They may provide an accurate portrayal of a particular system, for example, winter wheat growth in Nebraska, but are not transferable to other situations such as winter wheat growth in eastern Russia. System simulation models tend to be quite realistic, somewhat precise, but not very general. Control models take up various positions in the figure, depending on the level of precision desired.

B. Aggregation and Loss of Detail

A major factor that affects where a particular model fits into the scheme shown in Fig. 3 is the amount of aggregation included. Natural systems consist of many components that can be lumped together or disaggregated. Population models that use a single variable to represent the whole population must inherently ignore the within-population structure (e.g., age and size). Such a model would not be able to discriminate between a population with mostly small individuals and one with mostly large individuals, unless this within-population structure was assumed to affect the population's growth characteristics in some manner. Aggregation implies a loss of detail that allows for quite general models. Often this gives model parameters that are not at all easily estimated from observations. Examples would be the growth rates in population models and the competition coefficients in community models. The less aggregation in the model, the more parameters there tend to be and more data are required in order to estimate them.

C. Uncertainties: Mechanisms, Data for Parameterization, Biotic and Abiotic Forcing Functions

The modeling process is limited by the information available. There may not be basic agreement on the mechanisms that are critical in the system of concern, so that any particular model includes only some of the mechanisms, or just partial information on these mechanisms. Many population models applied to vertebrates ignore social structure, despite evidence that it is often present in such populations. This is typically due to lack of understanding of the effect of such structure at population levels, and gives rise to another set of models designed specifically to investigate these effects.

Under situations in which the mechanisms are well understood, it may not be possible to accurately estimate model parameters because adequate data are lacking. There are a variety of statistical methods designed specifically to determine the optimal choice of parameters in such situations (Hilborn and Mangel, 1997). These methods may take account of parameters estimated either for similar species or for similar locations other than the one being considered. For example, it may be necessary to use observed clutch size distributions for one bird species in application to a similar species about which such information is lacking. Another uncertainty associated with natural system models arises due to the unpredictability of forcing functions such as weather and disturbance. If historical information is available on these functions, this may be used to estimate the stochastic effects of such forcing. These various uncertainties limit the detail at which models may be constructed, and thus limit the types of questions that may be addressed using models.

V. SOME TOOLS OF THE TRADE

A. Statistical Approaches

Statistical models usually have a descriptive objective rather than a mechanistic one. The parameters within these models are directly estimated by choosing them in a manner that best fits a certain data set. Thus any particular statistical model is typically not very general in application to different systems. The structure of such models may be useful in a wide variety of different contexts. Regression models, which assume a particular mathematical relationship between variables and assume that errors in the data take a particular form, are widely applied. Numerous regressions have been estimated for species richness as a function of latitude, altitude, and rainfall (Huston, 1994). Discussion of statistical methods applied to estimation of population sizes and densities may be found elsewhere in these volumes.

B. Dynamic Models

Although many of the traditional models for populations and communities are in the form of dynamical systems (e.g., collections of linked differential or difference equations), often the types of analyses performed for these models are based around equilibrium assumptions. The objective is to find long-term asymptotic behavior. This may be a static equilibrium (e.g., population sizes approach a constant value through time) or a dynamic equilibrium (e.g., population sizes follow repeatable patterns through time) (Murray, 1989). While these situations may arise, many models produce behavior that does not have a long-term equilibrium structure. Another key objective is to determine the stability characteristics of any equilibria that arise, in the mathematical sense of determining whether a model that is perturbed from an equilibrium condition will return to it. The dynamics arising in all these models takes account of the basic demographics of the population, as well as interactions with other populations. Adding abiotic conditions such as temperature and rainfall, or adding spatial components, often requires that the analysis be done using numerical simulations.

C. Geographical Information Systems

The advent of remote-sensing methods using airplane cameras and satellite imaging has opened new possibilities for following and modeling the responses of the earth's biota. A key tool that allows the use of such materials are Geographical Information Systems (GIS), which enable computers to graphically display the remote-sensing data as two-dimensional maps. Each image may represent one aspect of an underlying landscape, such as landcover or vegetation type. The image value at any particular location (or pixel) in the map is estimated using models that classify the output of the cameras or the multispectral scanners on satellites into types appropriate for the objective. These models require ground-truthing to ensure that the estimated value for a particular location matches what is actually present. GIS methods allow various spatially explicit components of a landscape to be combined by looking at different map layers (different images measuring different aspects of the landscape). A mathematical function is then applied that averages or applies thresholds to these various components. Estimates of regional and worldwide carbon uptake are obtained using such methods applied to vegetation maps, in which different carbon assimilation values are assigned to different vegetation types, linked with weather maps supplying temperature and rainfall patterns.

VI. APPLICATIONS

A. Habitat Suitability Indices

Habitat Suitability Evaluation Procedures (HEP) are a formalized methodology for impact assessment on wildlife habitat. These are based on Habitat Suitability Index (HSI) models, which attempt to summarize the site characteristics that affect the utilization of particular habitats by a variety of wildlife species. Numerous HSI models have been constructed, typically consisting of very simple regression-type models. The key habitat variables are often some measure of canopy cover in a variety of classes, diameter classes of trees and shrubs, tree stem densities, area of open water, and distance to forest cover, among others. The objective is to combine these variables, based on extensive field observations done in a correlative manner, to provide overall indices of suitability. HSIs are always indexes with values between zero and one, and they are assumed to be proportional to carrying capacity.

HSIs are based only on local habitat variables; they completely ignore any effects due to species interactions, except those due to indirect effects on related habitat variables. The models ignore the spatial interactions of habitat types across a landscape. This leads to difficulty in situations for which the size, shapes, edge effects, and neighborhood relationships have a greater effect on habitat preference than local forest composition and structure variables. The models also do not take account of the issue of presence/absence of a species, and thus ignore any historical influence on potential local abundance. HSIs are inherently static entities, so any dynamics they produce are driven completely by changes in habitat variables and not by the inherent dynamics and demography in the species being considered. Despite these criticisms, HSIs are perhaps the most commonly used set of ecological models, in part because users realize their limitations and view them as a simplified tool to summarize a very complicated situation by a single number. Such simplification must result in a loss of information, but a key issue regarding HSIs is whether they indeed can be used as a useful predictor for abundance. A summary of the HSI modeling approach is given in Verner *et al.* (1986).

B. GAP Analysis

The Gap Analysis Program (GAP) is a nationwide comprehensive effort in the United States to inventory plant and animal species, computerize the results, and have the capability to analyze spatially the relationships among different taxa. GAP relies heavily on GIS methods, and has as a major objective the capacity to identify "gaps" in biodiversity. Such gaps are presumed to arise in locations that are expected, from biological knowledge of species requirements, to have certain mixtures of species present, but for some reason do not. The focus of these efforts is not on rare or threatened and endangered species, but rather on the status of ordinary species and their habitats in order to inform policymakers to improve their decisions. Much of the remote-sensing information cannot determine the presence or absence of particular species, because the technology can provide information only on basic vegetation components of a landscape, and landforms such as rivers, roads, and urban areas. Thus, an extensive ground effort is under way to survey the distribution of various species, map these, and relate these mapped distributions to habitat variables, land use, and ownership, and other related species distributions. The focus is on vertebrate species, basic floristic types whose presence can be estimated from satellite image analysis, and landform information.

Gap analysis does not attempt to explicitly model biodiversity, but rather uses the preceding surrogate measures as a method to assess spatial patterns in biodiversity. There are acknowledged taxonomic biases in this approach. The objective has been to provide basic methods to assess the impacts of the rapid land use changes that have occurred in the United States over the past several decades and how these impact biodiversity. Other limitations of this approach are the lack of detailed demographic information included for the species of concern, the static nature of the project, and the reliance on mapping methodologies that require extensive ground-truthing to ensure accuracy. Full documentation on the GAP may be found on the web at http://www.gap.uidaho.edu/gap. A basic reference is Scott *et al.* (1993).

C. Metapopulation Models

Both HEP and GAP are attempts to account for the spatial nature of biodiversity by including explicit maps of basic habitat variables. An intermediate approach between models that include all the spatial detail available from maps and those that ignore all spatial aspects of a system are metapopulation models. These models consider a landscape to be split up into a number of localized populations, called subpopulations, with the entire collection of these called a metapopulation (Gilpin and Hanski, 1997). Most of the biotic interactions driving population dynamics occur within the localized subpopulations, but there are exchanges of individuals between these subpopulations. This allows for differing environmental, demographic, genetic, and disease situations to be present in the subpopulations. Depending on the assumptions about transport of individuals between the subpopulations, these can be relatively isolated or closely coupled.

A clear advantage of metapopulation models is the ability to derive analytic results, such as equilibrium and stability behavior, as a function of the within-subpopulation characteristics and the between-subpopulation factors such as movement. The models are particularly appropriate for cases in which a landscape can be reasonably viewed as containing discrete patches of habitat suitable for the population, with the intervening regions not suitable. The level of detail in these models can be quite variable, with the simplest versions just treating subpopulations as either present of absent. More complicated models take account of demographics within each subpopulation or explicit details on the relative spatial locations of each subpopulation that affect movement between them. These can be used for a population viability analysis, in which the probability that the overall population will survive for varying time periods is estimated.

D. Individual-Based Approaches

All of the foregoing approaches include various levels of aggregation in their components in order to simplify the model. An alternative reductionist approach is to take account of differences between individuals within a population, allow the individuals to feed, grow, and interact, and from the aggregated behavior of these individuals build an understanding of population-level responses (DeAngelis and Gross, 1992). These individual-based approaches are increasingly common thanks to advances in computer science and the increasing availability of data on behavior and physiology of species of particular interest. The advantages of these approaches include the ability to consider the effect of abiotic factors on populations through their direct impact on individual behavior and growth, to take account explicitly of spatial variation in habitat factors, and to deal with small populations in which individual differences within the population can have great impacts on population-level responses. Disadvantages of individual-based approaches include the requirement for a great amount of detailed data to realistically simulate individual behaviors, and the typical necessity of making numerous simulations to evaluate any particular response that may arise because of the stochastic nature of the models.

VII. FUTURE TRENDS

A. Multimodeling and Regional Assessment

Natural systems have many interacting components operating at a variety of temporal and spatial extents and requiring differing levels of detail to describe the interactions between them. One historical approach in ecology to model such systems is to break it into a number of compartments (often for different trophic levels, and sometimes with grouping within each trophic level) and consider the dynamics of each compartment with movements of energy, biomass, or nutrients among them. It is quite difficult to make these systems analysis approaches spatially explicit or to link them to GIS. Yet it is now becoming possible to link together a variety of different modeling approaches in order to best utilize the available data, with different resolutions at different trophic levels, and carry this out in a spatially explicit manner. Such multimodels may use very simple models similar to HSIs for some trophic components, more complex dynamical systems for certain populations with mostly very localized interactions, and individual-based models for organisms that move great distances and average over the spatial heterogeneity. One example of such an approach is the ATLSS (Across Trophic Level System Simulation) Project, which is an ongoing attempt to build a multimodel for estimating the biotic impacts of alternative water management plans on the Everglades of south Florida (DeAngelis *et al.*, 1998).

Building multimodels requires extensive landscape data obtainable from remote-sensing and ground efforts. The approach is inherently dynamic, and thus requires methods to estimate the spatial dynamics of key environmental drivers, or else have available a history of this spatially that can be analyzed statistically. In the Everglades, the major driver is water, and both historical data and detailed hydrologic models are available to provide estimates for scenario evaluations. Without such data, assumptions must be made about the dynamics of the landscape. With the reduced cost of data storage and the development of standards for spatially explicit data, it is expected that future models will have available extensive time series of remotely sensed images to both calibrate multimodels and provide the opportunity to iteratively improve their predictive abilities. For problems at regional levels, such multimodels are a rational method to aid planning while taking account of the best scientific data at the variety of resolutions available.

B. Behavioral Dynamics of Species of Special Concern

Great strides are being made in improving our understanding of the conservation biology of rare, threatened, and endangered species throughout the world. In an effort to better estimate the responses of populations of these species, numerous remote-sensing methods have been employed to track the movements of individual organisms. These include radio-tracking devices implanted within sampled individuals, which allow explicit location and physiological data (e.g., body temperature) to be obtained regularly throughout the individual's life from either satellites, airplanes, or sampling locations on the ground. The technology has advanced so that implantation of such devices can be done rapidly with no determinable harm to the individual. When done for many individuals within a population, it is becoming possible to follow the behavioral dynamics of mixtures of individuals. This includes the ability to ascertain details of mating, territoriality, and aggressive interactions.

It is likely in the future that managers concerned with a particular species will be able to observe in real time the movements of many individuals within a population. Then they might apply spatially explicit modeling methods to project the response of the population to particular management alternatives. This scenario suggests that the simultaneous monitoring of other species that interact with the species of concern would allow for the regular application of adaptive management methods.

Acknowledgments

The author appreciates the support of the U.S. Geological Survey, Biological Resources Division, through Cooperative Agreement No. 1445-CA09-95-0094 with The University of Tennessee. The statements, findings, conclusions, recommendations, and other data in this report are solely those of the author and do not necessarily reflect the views of the U.S. Geological Survey.

See Also the Following Articles

ECOSYSTEM FUNCTION MEASUREMENT • MEASUREMENT AND ANALYSIS OF BIODIVERSITY • REMOTE SENSING AND IMAGE PROCESSING.

Bibliography

Caswell, H. (1989). *Matrix Population Models*. Sinauer Associates, Sunderland, Massachusetts.

Comiskey, E. J., L. J. Gross, D. M. Fleming, M. A. Huston, O. L. Bass, H.-K. Luh, and Y. Wu. (1997). A spatially-explicit individual-based simulation model for Florida panther and white-tailed deer in the Everglades and Big Cypress landscapes. In *Proceedings of the Florida Panther Conference, Ft. Myers, Florida, Nov. 1–3, 1994* (D. Jordan, ed.), pp. 494–503. U.S. Fish and Wildlife Service, Washington, D.C.

DeAngelis, D. L., and L. J. Gross (eds.). (1992). *Individual-Based Models and Approaches in Ecology*. Routledge/Chapman and Hall, New York.

DeAngelis, D. A., L. J. Gross, M. A. Huston, W. F. Wolff, D. M. Fleming, E. J. Comiskey, and S. M. Sylvester (1998). Landscape modeling for Everglades ecosystem restoration. *Ecosystems* **1**, 64–75.

Gilpin, M. E., and I. Hanski (eds.). (1997). *Metapopulation Biology: Ecology, Genetics and Evolution*. Academic Press, San Diego, California.

Haefner, J. W. (1996). *Modeling Biological Systems: Principles and Applications*. Chapman and Hall, New York.

Hilborn, R., and M. Mangel. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, New Jersey.

Huston, M. A. (1994). *Biological Diversity*. Cambridge University Press, Cambridge, United Kingdom.

Jambu, M. (1991). *Exploratory and Multivariate Data Analysis*. Academic Press, Boston.

Langton, C. (ed.). (1988). *Artificial Life*. Addison–Wesley, Redwood City, California.

Levins, R. (1968). *Evolution in Changing Environments*. Princeton University Press, Priceton, New Jersey.

Murray, J. D. (1989). *Mathematical Biology*. Springer-Verlag, Berlin.

Scott, J. M., et al. (1993). *Gap Analysis: A Geographic Approach to Protection of Biological Diversity*. Wildlife Monographs No. 123, Supplement to *The Journal of Wildlife Management*, Vol. 57, No. 1, January 1993.

Verner, J., M. L. Morrison, and C. J. Ralph (eds.). (1986). *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates*. University of Wisconsin Press, Madison.

CONSERVATION BIOLOGY, DISCIPLINE OF

Andrew P. Dobson* and Jon Paul Rodríguez[†]
**Princeton University and [†]Centro de Ecología, Instituto Venezolano de Investigaciones Científicas*

GLOSSARY

Allee effect Biological phenomenon whereby the survival of the individuals of a population increases by spatial aggregation.

community All forms of life that coexists and interact with each other in a particular habitat.

ecosystem Complex dynamic system that arises by the integration of a biological community with the physical environment.

***ex situ* conservation** Maintenance of live components of biodiversity physically removed from their original habitat.

extinction vortex Positive feedback loops that increase the risk of extinction of a population as it declines in size.

gene Functional unit of heredity; the portion of a chromosome of an organism that encodes for a particular protein.

gene bank Facility where genetic material is stored in the form of seeds, pollen, embryos, or semen, or, in some cases, as live plants or animals living in a field, a greenhouse or other installation.

habitat corridors or connections Strips of habitat that connect isolated habitat patches in a landscape transformed by human land use. Connections can be achieved through the conservation of existing habitat or by ecological restoration.

habitat restoration Active modification of the current state of a degraded habitat in order to return it to a former, preferred state.

***in situ* conservation** Maintenance of biodiversity within the ecological and evolutionary dynamics of the ecosystem where it exists.

population A group of individuals with common ancestry that are much more likely to mate with one another than with individuals from another such group.

CONSERVATION BIOLOGY is one of the fastest-growing fields of modern scientific research. It is an applied discipline that integrates principles of natural and social sciences with the objective of achieving the long-term persistence of biodiversity on Earth. This article reviews current trends in conservation biology, using a hierarchical organization to present the principal domains of research. These domains range from the largest level of ecosystems and communities, to the intermediate level of species and populations, and down to the smallest level, that of individuals and genes.

I. DEFINITION

The current global crisis in the loss of biodiversity is the result of the immense success of one species, *Homo sapiens,* at the expense of the majority of most other species. Concern about the loss of biodiversity arises from spiritual, moral, and aesthetic motives, through economic rationales, to purely selfish reasons. Moreover, the growth of these concerns has fueled the rapid expansion and awareness of conservation biology into academia, government, industry, and other sectors of society worldwide. Although the first conservation biology textbooks and scientific journals date back to the early 1960s, conservation biology's growth as a scientific discipline during the last two decades has been rapid and global.

Conservation biologists have an enormous agenda. Initially, they sought to understand the impact on biodiversity of the complex interactions between human societies and the natural world. They hope this understanding will provide guidelines for minimizing the negative effects of human actions on the persistence of biodiversity. Increasingly, the discipline has grown to recognize the great extent to which humans depend on biodiversity, and has begun to explore the services that natural ecosystems provide to humans. Ultimately, conservation biology aims to guarantee the persistence of natural landscapes, healthy ecosystems, species, populations, and genes, as well as the complex interactions among them and the dynamic processes that characterize them. This pursuit is necessarily trans-disciplinary, integrating the principles and practices of a variety of disciplines, principally ecology, population genetics, economics, and biogeography, but also increasingly anthropology, psychology, engineering, law, and public policy.

II. ECOSYSTEMS AND COMMUNITIES

The mandate of conservation biology covers a spectrum that runs from concern about the conservation of large, intact, functioning ecosystems, through the maintenance of viable indigenous human and ecological communities, to the preservation of the last surviving individuals of species in zoos and botanical gardens. Arguably, this spectrum may be extended to include the deoxyribonucleic acid (DNA) of currently extant or recently extinct species in museums. This spectrum includes both processes that are complex and only partially understood and ones that are more completely characterized.

The most economical and effective way to conserve biodiversity is through *in situ* protection that guarantees the persistence of all the processes that characterize biodiversity in a fully functioning ecosystem. Biogeochemical and hydrologic cycles, the flow of energy through food webs, and evolution by natural selection all require a complex web of interacting biological and physical elements. It is in general unrealistic to assume that these organisms and the processes they mediate can be removed from the natural setting and recreated in a zoo, botanical garden, theme park, or gene bank—for long-term persistence, they must be conserved where they stand.

The most popular instruments for *in situ* conservation are nature reserves and protected areas. According to the World Conservation Union (best known by its acronym, IUCN), a protected area is a section of land and/or sea especially dedicated to the protection and maintenance of biodiversity and of natural and associated cultural resources, which is managed through legal or other effective means. Protected areas come in many kinds and differ mainly in size and the degree of human intervention that is allowed to occur within them. At one extreme are large, strict nature reserves and wilderness areas, where little or no obvious modification of the landscape is permitted. Examples of these include the Bob Marshall wilderness in Montana, United States, and the Parima-Tapirapecó national park in southern Venezuela. At the other extreme are "managed resource" protected areas, which are devoted to the sustainable use of natural ecosystems, such as the networks of game reserves in Botswana or Spain. In all cases, the maintenance of ecological and evolutionary processes is recognized as a key motivation for the creation of such reserves. Ecosystem studies play an increasingly major role in understanding complex ecological systems, and they provide important guidelines for the management of protected areas.

Although there are very few examples of nature reserve systems that were based from their inception on ecological principles, conservation biologists have undertaken many post hoc analyses of reserve systems that were originally designated for political, aesthetic, and occasionally scientific reasons. For example, Yellowstone National Park in the United States, the world's first national park, was founded "For the benefit of the People" so that they could appreciate its spectacular and unique geohydrothermic features. By chance, it also happens to contain a nearly complete set of the fauna and flora found in the North American Northern Rockies, and thus is also valuable from the perspective of the conservation biologist.

The first systematic approaches to examining the design of nature reserve systems originated in the application of R. H. MacArthur and E. O. Wilson's "theory of island biogeography" to fragmented landscapes. This theory postulates that the number of species on islands is the result of a dynamic equilibrium between the processes of extinction and immigration, such that as island size increases, extinction rate decreases, and as distance from the mainland increases, immigration rate decreases. As a result, a small island located far away from the mainland will achieve a smaller equilibrium number of species than a large island located near the mainland (Fig. 1). This led Jared Diamond to suggest a series of geometric arrangements that compared preferred to less optimal distributions of patches of protected habitat (Fig. 2).

The creation of a nature reserve may reduce or completely halt habitat conversion within its borders, but it not likely to prevent land use changes in surrounding areas. As time passes, nature reserves become habitat "islands," and may effectively decrease in size while increasing in degree of isolation from the matrix of habitat that once included them. The theory of island biogeography predicts that the equilibrium number of species in habitat islands should decrease until either the effective size of the nature reserve stops decreasing or the effective distance from sources of immigrants stops increasing. Thus the effectiveness of many parks in the long-term conservation of biodiversity is compromised by processes that erode their boundaries and by those that increase their isolation.

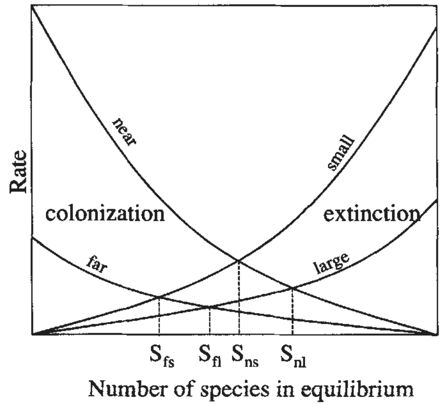


FIGURE 1 The theory of island biogeography predicts that the number of species in equilibrium for a small island located far away from the mainland is smaller than for a large island located near the mainland. [Redrawn from B. A. Wilcox, Insular ecology and conservation, in *Conservation Biology: An Evolutionary-Ecological Perspective* (M. E. Soulé and B. A. Wilcox, eds.), pp. 95–118, © 1980, with permission from the editor.]

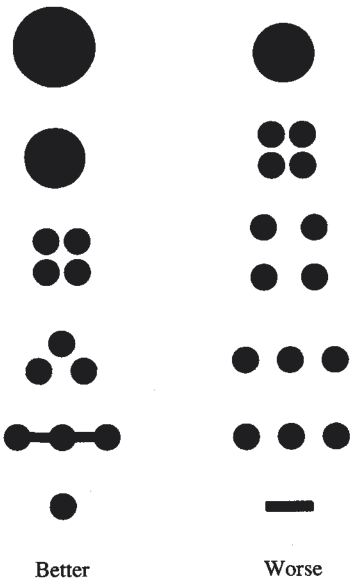


FIGURE 2 According to the theory of island biogeography, large, round, connected, or closely located protected area networks are better for assuring the persistence of a larger proportion of the species present in an area. (Reprinted from Biological Conservation, Vol. 7, J. M. Diamond, The island dilemma: Lessons of modern biogeographic studies for the design of natural reserves, pp. 129–146, © 1975, with permission from Elsevier Science.)

A. The SLOSS Debate

With time, conservation biologists have become aware that the theory of island biogeography might offer additional insights to questions related to the design of nature reserve networks. For example, what sort of actions would the theory prescribe to minimize the rates of species loss? Alternatively, how can a reserve network maximize the number of species preserved? Is there an optimal shape for nature reserves, or can they be arranged in optimal geographic designs?

In attempting to answer some of these questions, conservation biologists initiated a controversy that came to be known as the *single large or several small—or SLOSS*—debate. The main point of contention was whether the conservation of biodiversity in a particular area was better served by one large reserve or by several small ones. Proponents of the "single large" approach argued that large reserves were able to house large population sizes, thus minimizing the risk of extinction of the species present. In small reserves, high extinction rates would result in changes of the species assemblages in the reserve and lead to habitats that were very different from those initially targeted for conservation. Proponents of the "several small" approach demonstrated that several small reserves whose combined size equaled that of one single large reserve could include a larger proportion of the species present in a given region. Furthermore, several small reserves generate a set of independent duplicate subpopulations, which decreases the likelihood of extinction of the population as a whole from diseases or catastrophes.

Spatial considerations also play a major role in the size of areas set aside for endangered species. Indeed, the problem of incomplete consideration of spatial scale was at the heart of much of the SLOSS debate. For example, consider a small town that decides to set aside an area of habitat as a nature reserve. At this spatial scale, there is little difference in the species composition of similar-sized patches; however, a relatively small increase in the size of the patch is likely to capture a significantly larger proportion of the local biodiversity. In contrast, as we increase the scale of parks to the state, national, or even semicontinental scale, similar small increases in the area preserved will not significantly affect the species composition in the patch—but similarity between patches of habitat of similar sizes will decline. At this scale, it will be better to purchase, set aside, or otherwise preserve a variety of different patches of habitat. Indeed, at the national level of reserve planning, it is highly unlikely that the full wealth of a nation's biodiversity could be captured within a single reserve, unless that reserve approaches the size of the whole country or continent. Thus, the answer to the SLOSS debate may depend on the spatial scale considered.

Spatial and temporal considerations of scale are also central to our understanding of how ecosystems respond to natural and anthropogenic disturbances. The relationship between the spatial scale of a disturbance, the frequency of a disturbance, and the length of time it takes to recover from a disturbance will determine how readily degraded land can be recovered and used to increase the effective size of a preserved area. In habitats that require disturbances such as fire to maintain all their natural components of biodiversity, such as the New Jersey pine barrens in the eastern United States, or the banksia woodlands in western Australia, successful management will have to mimic the scale and frequency of natural disturbances. Some preliminary synthesis suggests that the recovery time from disturbance scales roughly with the square root of the area disturbed.

B. Gap Analysis

Modern reserve design methods seek to maximize the representation of the components of biodiversity in protected area networks while minimizing the resources needed to do so. For this approach to be effective, comprehensive information is needed on the geographical distribution of species from a wide range of different taxa.

The first step in a reserve design process is to select a series of surrogate measures of biodiversity and assess the existing network of nature reserves for how well it samples this biodiversity. This allows managers and researchers to identify components of biodiversity that are unprotected or not well represented in the current reserve network. The spatial configuration of these missing components defines priorities for future protected area designation. This technique, known as *gap analysis,* was pioneered by Australian scientists in the mid-1970s and popularized by U.S. governmental agencies in the decades that followed.

Gap analyses typically focus on well-known species or groups of species, and assume that their presence in a nature reserve network can be used as an indicator of the presence of other components of biodiversity as well. The efficacy of gap analyses relies on considering as much information as possible, and updating these results as new information becomes available. Sources of information need not be limited to biological data. Gap analyses may be carried out on the basis of climatic zones, landscape properties, or any subset of features in a geographical region and for any proportion of population size or geographical extent. In fact, given that for most regions of the world knowledge of the detailed distribution and abundance of species is quite limited, the best available data may be a combination of abiotic environmental parameters that are known to correlate with the biological species of interest. These data can then be integrated into a program of biological surveys that constantly adds new information to the database.

Figure 3 depicts the steps in a gap analysis. This process frequently employs "remotely sensed" data obtained from satellites and requires the use of geographical information systems (GIS). Maps of the distribution and abundance of plants and animals, vegetation types, topography, climate, and soil properties (among other possible data sources) are combined to generate overlay maps that quantify landscape heterogeneity. A map that combines existing protected areas is contrasted with the target features overlay

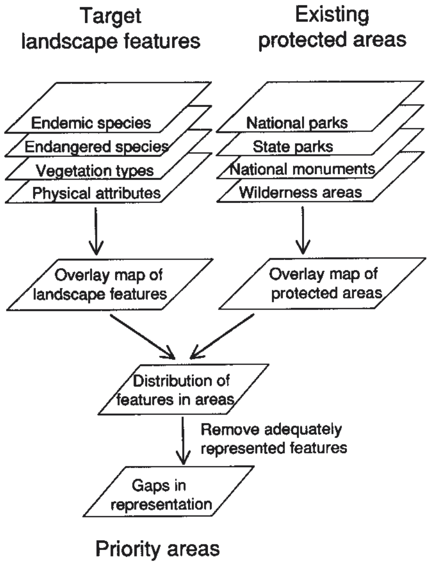


FIGURE 3 Schematic of the gap analysis process.

map and the presence of these attributes in the nature reserve system is assessed. Thus, a final map with the distribution of those attributes that are not included in nature reserves allows for the identification of gaps in protection.

The definition of what constitutes adequate representation is the next challenge. Assessors may be interested in assuring that all species, communities, or habitats considered important are present at least once in a network of protected areas. After achieving this most basic objective, other criteria may be added that minimize costs or future threats, or that maximize the total area of the reserve system by considering alternative configurations of land. It may also be possible to select alternative configurations to reduce conflicts with alternative land uses for the area under consideration.

An important problem that arises is that even if a species is present in a protected area at the time of designation, this does not assure that this species will persist in the long term. Furthermore, the presence of a set of target species does not assure that their parasites, predators, prey, or mutualists are also included, or that other unrelated species will be present as well. The protection of a target species may not provide an adequate measure of the status of ecosystem services, such as the prevention of erosion or the pollination of crops. However, these are not necessarily weaknesses of the gap analysis process itself, but of the way in which it is implemented.

C. Reserve Selection Algorithms

The costs of administering a reserve network increase proportionally to the area protected. Two fundamental concerns of protected area designers are to efficiently allocate available resources, and to minimize the area required to protect as many natural features as possible. The efficiency of a reserve network can be described by the efficiency (E) index

$$E = 1 - (x/t),$$

where x is the number or area of the sites required to achieve a particular representation target and t is the total number or area of sites to choose from. Thus, a network of protected areas is more efficient as the amount of land required to achieve the target decreases.

Reserve selection algorithms provide systematic means for designing efficient reserve networks. They follow a step-wise process for choosing a subset of sites from a list of proposed sites, such that the efficiency of the chosen subset is maximized. In other words, these algorithms allow for the selection of the minimum amount of land that is required to include all target species (or other attributes) at least once. There are three types of reserve selection algorithms. "Richness-based" algorithms begin the process by selecting the site containing the greatest number of species or attributes and sequentially adding those sites that include the greatest number of new species, until all species have been accounted for. "Rarity-based" algorithms select sites on the basis of rarity scores, favoring sites that are richest in species with restricted ranges. The third type of algorithm depends on linear-programming techniques, which sequentially search the database of available sites for replacement sites that increase the efficiency of the current set.

Although richness-based algorithms are the least efficient, they are the easiest to implement and have the advantage of including the principle of complementarity—that is, at each step of the process, the selected site is most complementary in terms of the features that it adds to the set. Rarity-based algorithms tend to be more efficient than richness-based algorithms; by focusing on species with restricted ranges, more wide-ranging species tend to be included by default, and fewer sites are required to achieve fairly complete levels of coverage. Though linear-programming methods are ultimately the most efficient, they rely heavily on intensive computing and are prohibitive for very large data sets.

Reserve selection algorithms allow for the selection of efficient reserve networks, but focus only on the adequate representation of a set of species or habitats within the landscape. They do not provide answers to questions related to the size, shape, or number of protected areas that are required to assure viable populations of different species, nor do they guarantee ecosystem functions and services. These more ambitious targets may only be achieved by expanding reserve networks as an interconnected system that allows the interaction of biological and physical components in a continuous landscape.

D. Corridors and Connectivity

Although overexploitation is responsible for the extinction of some species, the majority of species have declined because of the loss and fragmentation of their natural habitats, combined with the introduction of alien competitor, predator, pest, and pathogen species. It is these indirect effects of human development that have had the biggest impact on biodiversity. The crucial, and perhaps the only, way to minimize future impact is to reduce current and future rates of habitat loss and to explore ways of expanding current systems of reserves by the development of large-scale connections across the landscape.

Isolation of nature reserves and protected areas can often disrupt or break migration corridors that are used by seasonal residents of the park. The regular movements will range from the diurnal movement of snakes, birds, and many other organisms between feeding and nesting or brooding sites, to the larger annual migrations of large ungulates. In some instances, it is possible to maintain needed connectivity between habitat fragments by protecting natural corridors or stepping stones of appropriate habitat. In other cases, natural habitat and stopover points may be lost and enhancement of these converted habitats through ecological restoration may be the only option. The interactions between protected areas and developing or developed areas of human use outside them are often subtle and complex. Increased isolation is likely to lead to changes in the flow of nutrients and pollutants into and out of the protected area. Habitat conversion in the surrounding matrix facilitates the invasion of alien species that interfere with the species that the reserve was designed to protect.

Corridors and connections in the landscape have two major functions at the species level. First, they permit regular daily or seasonal movements, helping to ensure that different sub-sections of larger populations may have access to all the resources they require, while also maintaining the potential for all individuals in the population to successfully interbreed. Second, connectivity facilitates the dispersal of animals from their place of birth to their adult home range where they breed. At the regional or landscape level, this latter function is usually the most important, justifying, for example, the achievement of connectivity between mountain ranges in the western United States for the long-term conservation of fully functioning ecosystems.

The restoration of connectivity must occur at many scales. Most attention to date has been paid to local and regional connectivity as a way to redress the various forces threatening small populations. For example, amphibians must be able to move safely across a country lane during their annual migration to a breeding pond. However, many regional and inter-regional corridors have additional goals, such as accommodating the need for grizzly bears (*Ursus arctos*) to disperse safely between the Canadian Rockies and the Northern Rockies of the United States. In the face of global climate change and other major environmental changes, a substantial system of landscape connections is a major prerequisite for ensuring species persistence.

III. SPECIES AND POPULATIONS

Although conservation at the ecosystem level is the most effective way of conserving biodiversity, a significant proportion of research within the discipline of conservation biology has focused on conservation at the population level. This emphasis on species and populations partly reflects the constraints contained within legal mechanisms to preserve biodiversity; for example, both the Endangered Species Act in the United States and the global IUCN red lists focus on conserving species and/or populations. However, the species also offers a relatively well-defined (although not un-ambiguous) biological unit of study. Furthermore, the global biodiversity crisis is often most effectively communicated in terms of increasing extinction rates or numbers of threatened species. Therefore, awareness building and fund-raising tend to be more successful when focused on a species, rather than an ecosystem or a particular environmental problem.

It is often easier to raise funds or public support for charismatic species, such as elephants (*Loxodonta* sp.), chimpanzees (*Pan troglodytes*), or giant pandas (*Ailuropoda melanoleuca*), than for a generic habitat, such as a tropical or bamboo forest. More pragmatically, it is often possible to use the requirements of a single species to maintain the viable complexities of entire biological communities or ecosystems. By focusing on either a target species that occupies a keystone position in an ecosystem or a species that requires large tracts of land to persist (often referred to as an "umbrella" species), many other additional species and their natural habitats are also protected. This is effectively the strength of the "jeopardy amendment" in the U.S. Endangered Species Act, which provides legislation that is designed to conserve habitat in order to ensure the viability of one specific species—as occurred with the northern spotted owl (*Strix occidentalis caurina*) and "old-growth forest" in the north-western United States. Similar species-centered approaches led to the creation of the Cockscomb Basin jaguar (*Panthera onca*) preserve in Belize, the Arabian oryx (*Oryx leucoryx*) sanctuary in Oman, the Laguna Brava vicuña (*Vicugna vicugna*) reserve in Argentina, the Wolong and Wanglang nature reserves for panda habitat protection in China, and the significant expansion of the El Guácharo national park in Venezuela to assure feeding habitat for a large oilbird (*Steatornis caripensis*) colony.

Population viability analyses are a valuable tool for conservation biologists concerned with studying the long-term persistence of particular species or populations. In the early 1980s, Mark Shaffer introduced the concept of the minimum viable population (MVP), as the minimum size required for an isolated population to have a 99% chance of remaining extant for the following 1000 years in a particular habitat. Shaffer's work provided a quantitative framework for thinking about species conservation objectives and the timescales within which human actions should be considered. While focusing on the MVP concept, researchers began to realize that the key to conservation at the population level was understanding the impact of different stochastic and deterministic processes that determine extinction risk. By examining the sensitivity of the population to changes in these processes, it was argued that management actions could be taken to increase population persistence. Although there was an unfortunate tendency for many early analyses of viability to emphasize captive breeding as a means of enhancing persistence, this is now perceived as an option of last resort. Moreover, the notion of an MVP was progressively assimilated into a series of techniques that are known as "population viability analysis."

A. Population Viability Analysis

The central theme of population ecology is represented in a large canon of work that examines how populations respond to intrinsic and extrinsic factors and how these affect their long-term dynamics and persistence. Although deterministic, stochastic, environmental, demographic, and genetic processes can all have complex and interacting effects on extinction risk, these processes tend to operate in a hierarchical fashion determined by the size of the remnant population. Furthermore, habitat destruction and overexploitation can exacerbate intrinsic risk factors and accelerate population declines. A population viability analysis (PVA) explores the interactions between different known factors and assesses the likelihood that a population will become extinct within a specified time frame and under particular circumstances. PVAs generally require the use of computer simulations and simple mathematical models. The work of M. A. Burgman, S. Ferson, and H. R. Akçakaya (1993) provides the definitive introduction to this area.

PVAs focus on individual species, but they can also examine the potential consequences of different ecological factors, including interactions with other species, rates of habitat conversion, or management interventions that may be taken to improve population persistence. When integrated sensibly into wildlife or endangered species monitoring programs, PVA can be a powerful tool to assist in decision making within an adaptive management framework. The degree of complexity of a PVA will depend on the number of elements that a manager decides to consider, yet more complex models do not necessarily mean more useful results. Understanding the model structure and its sensitivity to varying parameter values may be more important that the modeling process itself. Numerous PVA software packages are available. Unfortunately, the different assumptions hidden within different packages mean they can produce very different results from the same set of data. Thus, it is crucial to bear in mind that PVAs should not be used to estimate extinction risks per se. Instead, they are most useful for ranking different management options to determine those that increase the likelihood of persistence and that can be implemented at lower costs. They are also invaluable for indicating areas of ignorance about a species and thus focusing future research. The precautionary principle indicates that ignorance should always imply increased protection until a more complete set of demographic data on the species is available.

B. The Demography of Small and Declining Populations

The recognition that current losses of biodiversity constitute a global crisis has led to a growing interest in the problems faced by small populations. By definition, all populations that go extinct first become small. Research in this area has emphasized the positive feedback between a number of mechanisms, each capable of increasing extinction risk as populations further decrease in size. The four principal mechanisms are (1) chance demographic events, (2) environmental stochasticity, (3) inbreeding depression, and (4) Allee effects due principally to the breakdown of the social benefits accrued by living in larger units. Unfortunately, in the 1980s this "cult of the extinction vortex" led to an emphasis on crisis management, while relatively less attention was focused on understanding the processes that initially caused the decline in a previously viable population. In an insightful review of population-level approaches to conservation biology, Graeme Caughley (1994) proposed that the discipline has developed along two relatively independent but complementary lines: the "small population paradigm," which has received more emphasis and addresses problems related to the persistence of small populations, and the "declining population paradigm," which considers the causes of population declines and how to reverse them.

In broad terms, "small population" conservation biologists have focused on finding ways of increasing the size of small populations and achieving the maintenance of a regular flow of individuals between fragmented populations. The application of these principles to the conservation of endangered species in the wild has been limited, but they have played a major role in influencing captive breeding practices, while providing some insights into the theory of how nature reserve systems might be designed.

In contrast, an increasing number of conservation biologists have focused their research on identifying the causes that lead to declines in species abundance or the contraction of the range of a species. The primary forces operating here are habitat loss and fragmentation, although threats posed by alien species, overexploitation, and pollution are also important. By examining populations at an earlier stage in their decline, it is hoped that strategies may be found for halting and reversing the decline. This paradigm is concerned with prescribing early, preventative actions that require minimal intervention that may be taken before a population becomes seriously endangered. At present, this approach has not succeeded in generating elegant theoretical principles to guide management decisions. This is mainly because population declines can be caused by numerous factors operating simultaneously on a community of populations. Not surprisingly, these effects are difficult to quantify. The key to future success in this area is carefully designed experimental and monitoring programs that systematically seek causes for population declines.

The most likely future for these two paradigms is eventual fusion. Clearly, both approaches are useful and contribute complementary insights to the advancement of the discipline.

IV. INDIVIDUALS AND GENES

All forms of life carry genetic information. In the long term, the capacity of any species to adapt to anthropogenic and nonanthropogenic changes in the environment will be related to the existence of genetic variation. *In situ* conservation programs are almost without exception the best option for conserving the genetic diversity that is a component of biodiversity. The protection of an ecosystem and its ecological processes provides the necessary conditions for evolution and natural selection to proceed. Additionally, protecting species in the wild is much more cost-effective than maintaining them in captivity. In the cases of some species, however, such as varieties of domesticated plants and animals and highly endangered species, *ex situ* conservation technologies are required. In these cases, the maintenance of genetic variation becomes a central topic of concern, especially if an endangered species is taken into captivity for future reintroduction into the wild.

A. Measuring Genetic Variability

Genetic information is encoded in genes, which are mainly composed of long chains of deoxyribonucleic acid, or DNA. Genes aggregate in the cells of every organism to form chromosomes. Variation in genetic information can be measured at any step of the biological process that transforms the sequence encoded in genes into its protein products. This information can then be used to quantify differences between organisms, populations, species, or any other taxa.

Methods for measuring genetic diversity focus on the variation of "markers," much in the way that gap analysis uses species or topographic variables to characterize the degree of spatial heterogeneity in a landscape. These markers are assumed to correlate in a broad sense with the genome as a whole; the more variable a population is at marker loci, the more variable it is overall. In the past, scientists were limited to examining indirect or "phenotypic" measures of genetic diversity through markers such as variable external morphology or coloration, intracellular concentration of secondary compounds such as phenolics, cell pigments, or hydrocarbons, and variable protein products called "allozymes," which may be visualized using electrophoresis. In recent years, however, biologists have developed techniques for examining the DNA itself. This permits the use of markers such as the presence or frequency of specific base-pair sequences in DNA molecules. DNA fingerprinting, for example, allows the characterization of differences between individuals in a population on the basis of patterns in the variation of minisatellites, or relatively short (10–50 base-pairs long) repeat sequences that are distributed in tandem arrays throughout the nuclear genome. Variation in organelle DNA (such as mitochondria or chloroplasts) may also be examined. Different versions of these techniques are used to assay genetic differences within and between populations, species, or other taxonomic levels.

B. The Role of Zoos, Aquaria, Botanical Gardens, and Gene Banks

In extreme cases, such as when the last few remaining individuals of a species are under an imminent threat in the wild, or when a wild variety of a particular crop is being rapidly replaced by modern monocultures, the maintenance of populations in captivity may be the only short-term action to prevent their extinction. Zoos have played an important role in preventing the extinction of the Arabian oryx, the peregrine falcon (*Falco peregrinus*), and the Guam rail (*Rallus owstoni*), but in all cases the captive facilities were used to buy time while the agents of decline were being dealt with. Gene banks may be the last resource for hundreds of threatened varieties of domestic animals and plants. The Royal Botanical Gardens in Kew (United Kingdom) hold about 38,000 species in their collection; Brazil and Colombia are the only countries in the world that have more species within their territories.

The primary functions of these captive facilities are not to save endangered species, however, but to promote education and research. Conservation objectives are clearly most effectively and efficiently achieved by *in situ* rather than by *ex situ* approaches. Nonetheless, zoos, aquaria, and botanical gardens provide visitors with an opportunity for close interaction with a wide variety of plants and animals that are otherwise out of reach for most people. Zoos attract millions of people worldwide and are ideal facilities for educating them about the challenges and opportunities created by the present biodiversity crisis. They are also potentially important facilities for raising funds that can then be invested in conserving habitat in the regions where the animals and plants on display originally came from.

V. FUTURE CHALLENGES

The future of conservation biology lies in teamwork. The complexity of issues surrounding the conservation of biodiversity requires a trans-diciplinary approach that integrates knowledge and expertise from the different disciplines of natural and social sciences. Conservation biology will also become increasingly cross-institutional as governments, academia, and non-governmental organizations work together on the same issues but with complementary approaches and points of view.

An excellent model is provided by the Inter-governmental Panel on Climate Change (IPCC), a global organization dedicated to assessing the available scientific, technical, and socio-economic information in the field of climate change. The IPCC combines experts from around the world into three working groups that focus on (1) the climate system, (2) impact and response options, and (3) economic and social dimensions of climate change. Although the process is slow and long, the IPCC has already produced two assessment reports and is expected to produce a third report by the year 2000. The topics covered by the IPCC reports range from the state of the art of climate modeling to benefit/cost analyses of choosing between different mitigation options. The process has not been free of flaws and criticisms, but has allowed the identification of issues on which there is consensus. Likewise, it has allowed the identification of topics where agreement has not been reached, due to either a lack of scientific data or uncertainty about the nature of the processes underlying global change.

One requirement of conservation biology in the future will be the design of biodiversity management programs and decision-making mechanisms that incorporate, rather than reject, uncertainty. Although our understanding of ecosystem function and human dependence on nature's services has greatly increased, it is still far from perfect. As in the case of global climate change, we might never understand every detail of the numerous processes that influence biodiversity. This cannot remain an excuse for inaction, however. The adoption of an adaptive management framework that embraces the precautionary principle will allow us to use the best science available to decide on issues of biodiversity stewardship. One of the most difficult issues may be the revision of current resource consumption patterns, particularly in the developed world, and the design of new paradigms and technologies to attend to the needs of a growing global population. After particular actions are taken, their success may be monitored and adjusted in response to changing conditions, which arise because of either abrupt, unanticipated events or slowly evolving processes. With time, our understanding of biodiversity and our management decisions will both improve, generating creative solutions to reduce the impact of humans on Earth's systems while assuring the quality of life of all those who inhabit the planet.

See Also the Following Articles

CONSERVATION MOVEMENT, HISTORICAL • *Ex Situ, in Situ* CONSERVATION • ISLAND BIOGEOGRAPHY • KEYSTONE SPECIES • NATURAL RESERVES AND PRESERVES • POPULATION VIABILITY ANALYSIS (PVA) • RESTORATION OF BIODIVERSITY

Bibliography

Avise, J. C., and J. L. Hamrick. (eds.). (1996). *Conservation Genetics: Case Histories from Nature.* Chapman & Hall, New York.

Burgman, M. A., S. Ferson, and H. R. Akçakaya. (1993). *Risk Assessment in Conservation Biology*. Chapman & Hall, London.

Caughley, G. (1994). Directions in conservation biology. *J. Animal Ecol.* 63, 215–244.

Csuti, B., S. Polasky, P. H. Williams, R. L. Pressey, J. D. Camm, M. Kershaw, A. R. Kiester, B. Downs, R. Hamilton, M. Huso, and K. Sahr. (1997). A comparison of reserve selection algorithms using data on terrestrial vertebrates in Oregon. *Biol. Conservation* **80**, 83–97.

Dobson, A. P. (1995). *Conservation and Biodiversity*. Scientific American Library, Freeman, New York.

Margules, C., A. J. Higgs, and R. W. Rafe. (1982). Modern biogeographic theory: Are there any lessons for nature reserve design? *Biol. Conservation* 24, 115–128.

Meffe, G. K., and C. R. Carroll. (eds.). (1997). *Principles of Conservation Biology*. Sinauer Associates, Sunderland, Massachusetts.

Pressey, R. L., C. J. Humphries, C. R. Margules, R. I. Vane-Wright, and P. H. Williams. (1993). Beyond opportunism: Key principles for systematic reserve selection. *Trends in Ecol. Evol.* 8(4), 124–128.

Scott, J. M., F. Davis, B. Csuti, R. Noss, B. Butterfield, C. Groves, H. Anderson, S. Caicco, F. D'Erchia, T. C. Edwards, Jr., J. Ulliman, and R. G. Wright. (1993). Gap analysis: A geographic approach to protection of biological diversity. *Wildlife Monogr.* **123**, 1–41.

Soulé, M. E., and J. Terborgh. (eds.). (1999). *Continental Conservation: Scientific Foundations of Regional Reserve Networks.* Island Press, Washington, D.C.

United Nations Environment Programme (UNEP). (1995). *Global Biodiversity Assessment.* Published for UNEP by Cambridge University Press, Cambridge, United Kingdom.

CONSERVATION EFFORTS, CONTEMPORARY

Kristiina A. Vogt, Oswald J. Schmitz, Karen H. Beard, Jennifer L. O'Hara, and Michael G. Booth
Yale University

GLOSSARY

ecological integrity The balanced state of an ecosystem under normal environmental conditions, including the capacity of an ecosystem to absorb disruption and to recover from the disruption.

ecosystem engineers Animals (e.g., beavers) that modify an ecosystem by physically changing it (e.g., building dams) or by being present in such a great abundance (e.g., large grazing animals) that they have a disproportionately large impact on ecosystem functions.

ecosystem management Approach to management that attempts to maintain high ecosystem integrity while providing the services, uses, values, and products from that system for the long term. It explicitly integrates social, economic, and natural system sustainability.

invasive species Native and exotic species that are capable of displacing indigenous species or spreading into habitats where they were not common previously.

keystone species Species whose presence and abundance controls the integrity of a community or ecosystem and allows that system to persist within its natural range of environmental conditions.

legacy ecosystems Ecosystems that carry an intact "memory" of a change in some physical, chemical, or structural attribute or a change in species composition that modifies the resistance and resilience characteristics of the ecosystem.

resilience Rate at which an ecosystem is able to recover to conditions similar to those that existed prior to the imposition of a disturbance.

resistance Degree to which an ecosystem moves away from or is able to maintain current structure and function when faced with a disturbance.

restoration Re-establishment of similar structures and functions of an ecosystem or parts of an ecosystem that are no longer present because of past land uses or disturbances.

sustainability State that defines the biogeophysical, economic, social, cultural, and political thresholds in between which it is acceptable to continue to use and obtain services or products from a given piece of land. This definition requires all five factors to be considered when determining whether future generations will be able to acquire the same natural resources as the current generation.

sustainable development Meeting human needs without damaging the ecosystems that produces these resources and at the same time providing equitable distribution and access to these resources around the world.

system functions Processes that occur in an ecosystem that are typically measured as changes in the carbon or nutrient cycles regulated by decomposers, consumers, or primary producers.

umbrella species Species that have either large habitat needs or have other requirements whose conservation results in many other species being conserved at the ecosystem or landscape level.

THIS ARTICLE PRESENTS MANY OF THE ISSUES AND CHALLENGES currently being faced in converting theories from the natural sciences into practical applications for conservation. As part of this transition in conservation, two philosophical changes have created the most problems in making conservation practical. The first is a movement from a species to an ecosystem focus. The second change has been the recognition that humans influence the success of conservation projects and have to be factored into the conservation formula. The human influences have produced a dual approaches to conservation that varies between developed and developing countries. In developing countries, sustainable development is being linked to conservation. In developed countries, the conservation focus has been on restoration of species and habitats and in developing tools to eradicate invasive species. Four tools being used for managing and conserving species will be discussed: models, landscape patterns, reserve design, and economic evaluations.

I. INTRODUCTION: NEW CONCEPTS IN CONSERVATION

The discipline of conservation biology arose from the need to apply science to the protection of species that are threatened or on the brink of extinction. In its early years, biologists and ecologists had to worry about refining the science that would be brought to bear on the cause of protecting endangered species on particular tracts of land. This resulted in the practice of conservation being very species sensitive. Nowadays, conservation biology has to resolve many conflicting demands that are simultaneously being placed on ecosystems services that are provided by particular tracts of land, as discussed in the work of J. F. Franklin, J. C. Gordon, N. Myers, G. Daily, and others. Saving endangered species is typically only one component of a larger task to protect the entire biotic communities (biodiversity) that make up whole ecosystems. This has placed an emphasis on putting into practice many of the holistically based philosophical ideas that were previously articulated by John Muir, Aldo Leopold, and other visionary thinkers.

The mounting worldwide conflicts over the conservation of biodiversity and other uses of the land base make it imperative to develop and test new models of conservation. For example, many conservation efforts in the tropics have reinforced the importance of factoring in humans as integral elements of conservation for these efforts to be successful. As documented by the United Nations and presented by N. Myers, fully one-half of the world's biodiversity is locked up in tropical ecosystems where conflicts over resource uses are particularly intense. At the same time, human population size is at such a level in the tropics that the attendant population growth rates in these areas are double the world average. Such prodigious growth typically leads to a high level of poverty that makes conflicts over resource use and their protection particularly polarizing for resource managers and policymakers who are unable to satisfy all the stakeholders simultaneously. In these circumstances, even the best scientific knowledge about the ecology of groups of species will not provide the needed insight to protect biodiversity, since human demands on the natural resources will constrain the application of science to conservation. Therefore, the traditional model of conservation, whereby scientists continually refine and apply knowledge to the cause of species preservation, will be insufficient to meet the challenges of the future.

Additionally, the approaches to conservation have been evolving as new knowledge develops in population, community, and ecosystem and landscape ecology, as well as in social ecology. This knowledge has made us realize that the use of species demographic data is generally inadequate by itself for gauging conservation success and therefore for maintaining a species in the landscape. Species demographic data are effective in conservation when the species are keystone or umbrella species, or when they perform ecosystem engineering functions. Yet even then, knowledge of changes in population dynamics of a species does not necessarily provide a mechanistic understanding of the role that species play in ecosystems. Conservation has been moving towards holistic analyses of systems (i.e., ecosystem management) in which factors other than species biology dictate the approaches used to maintain species in their habitats. L. Harris, M. L. Hunter, M. Mangel, D. Simberloff, W. W. Weeks, and many others have documented this shift toward a holistic approach to conservation. Holistic analysis also recognizes that humans are an integral part of the system and cannot be considered as external factors in conservation projects.

It is clear that in order for conservation projects to be successful, sustainable economies must be created at the local level, as demonstrated by G. E. Machlis, J. McNeely, M. Wells, K. Brandon, and others. This recognition has resulted in the explicit linking of conservation with international sustainable development projects, especially in the humid tropics. Projects that meld these dual objectives are commonly called Integrated Conservation and Development Projects (ICDPs). These projects vary in scale and scope but typically includes one or more of the following topics as a central theme: biosphere reserves, eco-tourism ventures, non-timber forest product harvesting, and regional land use plans. Such projects are attempting to address the impacts of higher human population growth rates that are resulting in the development of economies that demand increasingly greater use of natural resources (e.g., tropical deforestation) to produce viable communities. The ultimate result of all of these changes is that humans are leaving a much larger "ecological footprint" on the land base and making it more difficult to practice conservation.

Conservation efforts also have to accept that species survival is heavily influenced by legacy ecosystems and landscapes. The end result of a legacy system is that it responds to future natural and human perturbations differently than do systems without these legacies, as shown by J. F. Franklin, J. J. Magnuson, T. A. Spies, P. M. Vitousek, K. A. Vogt, D. J. Vogt, and others. This is a problem for conservation since management and policy decisions may have to be based on science collected in non-legacy systems because that is the information that is available. Some factors contributing to the formation of legacy systems are:

- Past land uses and management practices that have fragmented the landscape and reduced habitat area, as documented by T. W. Clark, J. F. Franklin, R. T. T. Forman, and others.
- Introduction of invasive plant species that fix nitrogen, thus changing soil resource availability, as shown by P. M. Vitousek and, Walker in Hawaii.
- Existence of disturbances (i.e., hurricanes) that affect succession occurring at longer temporal frequencies than a manager's typical time frame (recently documented by many researchers working in Puerto Rico).

The insufficiency of habitat area for managing threatened grizzly bears has been well documented by T. W. Clark, S. C. Minta, and others for the Greater Yellowstone Ecosystem. In this case, the logical approach would be to restore a sufficient area of suitable habitat for the grizzly bear. However, it is insufficient to deal with just the habitat to ensure the survival of grizzly bears. Indirectly, grizzly bears are also threatened by an exotic fungus that is causing the decline of *Pinus albicaulis*, whose seeds are an important food sources for grizzlies. This underscores the need to recognize that conservation efforts are complicated by other factors regulating the functioning of a system that are not considered because they do not appear to have a direct bearing on a species.

Recently ecologists have come to understand the importance of disturbances that occur at long time frequencies (e.g., hurricanes) and how they affect the structure and drive the functioning of a system. Knowledge of these temporal dynamics is relevant for conservation, since the design of conservation efforts based only on consideration of intervals between disturbances can endanger an already threatened species. For example, when Hurricane Hugo swept over the tropical forests on Puerto Rico, the already small population of Puerto Rican parrots was dramatically reduced. These hurricanes are a reoccurring phenomenon there and have been recorded at a frequency of 21 years. The effect of these hurricanes on determining the structure and functioning of these forests during the intervals between hurricanes has been shown through the work of N. Brokaw, A. Covich, A. Lugo, F. N. Scatena, K. A. Vogt, D. Vogt, and others. Consequently, if ecosystem processes are studied only during the inter-disturbance intervals, we may overlook the important causal drivers for that system.

Five major philosophical changes are commonly identified when discussing why there is a need to adapt the old models and at times to find new models for conservation:

- First, there is the recognition that a focus only on species is inadequate to conserve species within habitats in perpetuity.
- Second, there is the realization that the human dimensional aspects of ecosystems cannot be ignored if conservation efforts are to succeed.
- Third, ecosystems are dynamic, so the maintenance of a particular type of habitat in the landscape will

require that several examples of that habitat exist elsewhere within that landscape but at different successional/developmental stages. This approach will require a landscape that is sufficiently large to include these stages. If the appropriate conservation area has been selected, when a particular habitat is lost at one location, its replacement will be developing at the same time in another part of the landscape.

- Fourth, conservation approaches will have to incorporate the tools of restoration ecology to reestablish the structural and functional aspects of native habitats required to maintain species. The restoration of damaged habitats may become a more central part of conservation practice, since centuries of human modification of the natural environment have resulted in inadequate habitat area or have sufficiently modified ecosystems. Because grizzly bear management in the Greater Yellowstone Ecosystem cannot be expanded beyond its present boundaries, there may be a need to consider restoration of habitats within the GYE if these habitats degrade due to the proliferation of invasive species. As the presence of invasive species increases within our landscapes, conservation biologists will have to consider using restoration tools as part of their management approach, since opportunistic invasive species frequently threaten the survival of native species.
- Fifth, disturbances (e.g., hurricanes and, droughts or floods associated with El Niño) at long timescales can eliminate or modify the habitat set aside in nature reserves so that the species being conserved are unable to survive in that diminished or degraded habitat.

These five changes in philosophy are causing a reassessment of how conservation biologists approach the practice of conservation. These changes are partially being driven by new scientific understandings that are beginning to filter into the mind-set of conservation biologists. In this article, we present a case for the need to increase the acceptance of these concepts by conservation biologists so that conservation practices can become more effective. These five major philosophical shifts have already resulted in several changes in the practices being used to implement conservation efforts (the most common changes are summarized in Table I). Two of these shifts will be discussed in greater detail in the next section. These two points have been selected for greater analysis because they have not been as extensively discussed in the literature, as have the other points (the reader is encouraged to consult the Bibliography for further readings on these topics). Section III will present a brief discussion of four important tools of conservation in the context of these philosophical changes.

II. ANALYSIS OF EVOLVING THINKING IN CONSERVATION BIOLOGY

A. Species to Ecosystem Shift

Because of the historical emphasis on saving threatened and endangered species, conservation biology has had a largely singular focus on protecting those species and their habitats. This focus was founded on a simple ethic that humankind has a moral obligation to protect all living things on earth. This species-centric approach seemed to be relevant in the early phase of conservation because anthropogenic changes to the land base were geographically localized and, by today's standards, comparatively benign. However, current approaches in conservation are strongly based on the ecosystem, landscape, and/or human system. This shift resulted because of the failure of many conservation efforts that were based on the protection of species. Yet it is important to recognize that the single-species approach should not be eliminated from the suite of tools available in conservation in the push to assess systems more holistically. It would benefit conservation if a framework was developed that helped managers identify which systems would be most likely to have single or multiple species playing critical ecosystem roles. A brief discussion of the types of single-species focus that are useful in conservation biology follows.

In some ecosystems, the single-species approach will continue to be a powerful tool for conservation biologists. For example, the single-species approach may still be appropriate as a primary conservation tool when a single species functions as a keystone species, as an ecosystem engineers, or as an umbrella species or coarse filters (discussed by W. J. Bond, C. G. Jones, R. T. Paine, M. Soulé, D. Wilcove, and others). In some ecosystems, keystone species can play an essential role in regulating ecosystem structure, ecosystem function, and/or the maintenance of community diversity. In the event that keystone species are lost from the system, the result may be a disproportionately larger effect on some property of the system so that certain species can no longer be maintained in the landscape. However, the concept of keystones may be difficult to apply in conservation because these species are not always obvious or we have not developed a clear mechanism for identifying them in the system, and some are only keystones part of the

TABLE I

A Summary of Some of the Dominant Practices Used to Implement Conservation Efforts in the Past and Today

	Previous practice	New practice
Conservation focus	Single species based approach. *Ex situ* conservation for single species (e.g., zoos, expensive reintroduction programs, captive breeding programs).	Umbrella species approach. Ecosystem based. *In situ* conservation model preferred. Restoration of species and ecosystems, elimination of invasive species explicitly linked to conservation.
Integration of humans and human impacts on ecosystems in conservation	Protected areas management or reserves could be designed to exclude humans. Conservation practices strongly based on managing spatial distribution of ecosystems or habitat scales in landscapes since the major human impact has been fragmentation and loss of ecosystems.	Conservation is human ecosystem based—explicit integration of people as primary drivers of the success of protected areas or reserve management and a need for people to access conservation areas. Recognition of the temporal scales of human land-use activities and how they produce legacy ecosystems and landscapes that control the success of protected areas or reserves needed in conservation (states that few non-human-impacted ecosystems exist).
Human values	Species values dominant and used to drive conservation practices.	Recognition that the species is one of the multiple values existing in a system and a species focus may miss the "real" driving variables determining whether an established protected area or reserve will be successful. Incorporation of and determining trade-offs of all values desired out of a land as part of conservation.
Disturbances	Incorporate short-frequency disturbances. Conservation plan attempting to exclude or limit human extraction of resources from nature reserves. Humans classified as disturbance agents who are negative to conservation efforts.	Inclusion of decade-scale disturbances (e.g., hurricanes) for species conservation. Conservation plan includes human extraction of resources from nature reserves. Explicitly integrates invasive species as a major disturbance factor affecting conservation success.
Spatial scale of analysis	Reserve design and size varied with territory needs of each species. Landscape matrix analysis recognizing influences outside of habitat fragment will impact conservation. Identification of edge and interior regions of reserves and their influence on species composition. Assumed ability to transfer information from plot to produce regional maps of endangered species. Use of corridors and buffer zones to link habitat fragments and reserve networks. Reserve design focus—single large or several small reserves, reserve size, where to place reserves.	Reserve design and size based on the ecosystem and landscape and not on species territory needs. Recognition that linking habitat fragments with corridors is inadequate. Knowledge that human land-use legacies and disturbance impacts modify quality of edge and interior habitats. Inability to transfer plot data to produce regional maps of endangered species unless characterizes ecosystems into their diverse types within a vegetative community.
Models	Small-scale, data-intensive species and community models.	Less data-intensive, large-scale biodiversity models and explicit linking to GIS. Ecosystem models geared to predicting human land-use activities and the sustainability of the ecosystems needed for conservation.
Economic evaluation	Development of nonmarket values for species.	Need to address all the values in each environment. Recognition that conservation efforts should not be driven just by economic analyses

time and in certain habitat conditions. Once keystone species have been identified, the ability to use the single-species approach simplifies conservation efforts. W. G. Conway suggested that the existence of keystones in an ecosystem would facilitate integrating conservation with restoration efforts. When a single species has not been identified that provides a key structural or functional role, species become poor indicators of whether a system is going to degrade or whether the habitat can be maintained for the species needing to be conserved.

The use of a single-species focus in conservation is greatly facilitated when species function as ecosystem engineers. In any ecosystem, the existence of ecosystem engineers is easy to recognize because they modify some structural part of the ecosystem that can be readily detected. They do not have to be present in large numbers for the impact to be registered. In many examples, a few individuals will have a disproportionately large impact on directly changing plant successional processes and indirectly modifying carbon and nutrient cycles (as has been documented for beavers and moose by R. J. Naiman, J. Pastor, and others). In addition, ecosystem engineers may play keystone roles at particular stages of the development of a vegetative community (i.e., their impact would not be measurable at all times).

In conservation biology, an important concept that has been discussed for more than a decade is the identification of single species that can function as an umbrella species or as a coarse filter (see P. F. Brussard, T. W. Clark, C. R. Tracy, and others). Umbrella species are those species that have large habitat needs (e.g., spotted owl, red-cockaded woodpecker, and grizzly bear) or have other requirements such that when the system is managed to conserve that species, many other species will consequently also be conserved. The single-species approach has been most often justified through the concept of umbrella species.

Unfortunately, many ecosystems do not have a single species that can be monitored to provide an indication of how that system is functioning and whether a conservation project will be successful. Based on our current scientific understanding, many species are merely passive inhabitants of ecosystem in which they live. Instead, entire communities of species may play integral roles in ecosystem function (several studies by J. J. Ewel, J. H. Lawton, S. McNaughton, S. Naeem, D. Tilman, and others have supported this idea). Biodiversity may then be a critical determinant of ecosystem productivity and stability—including the tendency to resist anthropogenic impacts and the ability to recover from impacts (i.e., resilience). However, the relationships between biodiversity and ecosystem productivity and/or stability cannot be generalized to all ecosystems. Strong relationships between these variables have been found in certain ecosystems (e.g., grasslands), but biodiversity cannot be automatically used as an indicator of ecosystem integrity (see K. H. Johnson, O. J. Schmitz, K. A. Vogt, and D. J. Vogt). For example, the biodiversity at the species level is a poor indicator of the resistance and resilience characteristics of woody perennial systems to disturbances since the response of the system occurs at levels other than species diversity (e.g., genetic diversity within a species).

The growing realization that biodiversity may be inextricably linked to ecosystem function has given rise to a more holistic, system-centric perspective for conservation. This new approach was codified by M. Mangel and others in 1996, when they defined a set of principles to which conservationists must adhere if we wish to successfully conserve wild living beings. These principles coincide fully with those articulated by J. K. Berry, K. A. Vogt, and others for successful ecosystem management. The principles of ecosystem management are:

1. focus on the sustainability of ecosystems, not on the output of products;
2. adopt a holistic understanding of the way all the parts are linked together in an ecosystem and the feedbacks among those linkages;
3. incorporate a long-term perspective and examine issues at a scale relevant to the functioning of the ecosystem; and
4. recognize that human values shape ecosystem structure and function in myriad ways that can constrain, promote, or reduce sustainability.

Because the primary goal of conservation is to sustain the habitat or ecosystem of species (as articulated by R. B. Primack), conservation biologists have already espoused the first principle of ecosystem management as articulated here. Most of the tools developed in conservation biology address this first principle but not the other three principles. Similar to what is still happening in ecosystem management (as discussed by J. C. Gordon, J. F. Franklin, K. A. Vogt, and others), conservation biologists are currently assessing the practicality of the different tools to implement the other three principles in conservation.

Out of the four principles listed here, the incorporation of people and their values into a holistic framework of analysis that is not solely human-oriented is one of the greatest challenges for both ecosystem management and conservation. Machlis has stated that conservation currently has no trained capacity to deal with the human dimensions of biodiversity conservation. Most of the dialogue related to integrating the social sciences into natural resource management is presently at the theoretical level (see J. K. Berry and K. A. Vogt).

Conservation is now facing challenges similar to those previously experienced by natural resource managers attempting to implement ecosystem management under conditions of high scientific uncertainty. For example, natural resource management shifted from a resource-based emphasis (e.g., forest stands, fish stocks) to a more holistic, ecosystem-based approach that better enfranchises all stakeholders of a particular land or resource base. In the face of such a shift, managers were forced to embrace new management models (i.e., adaptive management) to deal with large-scale management problems under a scenario of high scientific uncertainty about the response of ecosystems to human activities. Because our scientific knowledge base is still fairly limited, few guidelines exist to help resolve conflicts arising from the need to determine the trade-offs between different natural resource uses. These conflicts are especially difficult to resolve since conservation biologists and ecosystem managers have to deal with resource scarcities.

The current view of J. C. Gordon is that ecosystem and adaptive management are both approaches that are attempting to respond to resource scarcities. Ecosystem management is attempting to respond to a scarcity of land. This land scarcity means that the exclusive uses of any given land area for one purpose is no longer possible because the production capacities of ecosystems are finite. This contrasts with adaptive management, in which managers are responding to scarcity of scientific information and knowledge. This scarcity makes it difficult to develop consensus on which tools are most appropriate to use for resolving the conflicts over resource uses. In a similar vein, conservation biologists are being forced to respond to scarcity and the resulting conflicts that can arise over the use of finite resources. However, conservation has yet to adopt a model that can adequately handle conflicts resulting from the scarcity of resources and simultaneously achieve natural and social system sustainability.

B. Direct and Indirect Influences of Humans on Conservation Efforts

Humans are strong influences on whether conservation projects are successful. These human impacts can be understood by analyzing the different approaches taken by conservation organizations to address conservation issues in developed and developing countries. In developing countries, humans are dependent on natural resource extraction from the environment in which they inhabit. A major deterrent to conservation has been the inability of these governments to satisfy all the resource uses, services, and values that are expected from the same land base by local, regional, national, and international stakeholders. Under these circumstances, the scarcity of land area and resources means that conflicts arise over the uses of natural resources that typically polarize the people inhabiting the environment against conservation organizations articulating a need to establish reserves to save species. These conflicts often occur in environments where the land base constraints limit the options that are available to satisfy all the needs of all stakeholders. Some of the direct impacts of humans on the environment appear to be easier to manage, since alternative solutions are possible through substitution of products, technological breakthroughs that make resource extraction more efficient, shifting to the use of different resources, alternative employment opportunities, and so on. The issues that are relevant for developing countries will be further discussed in Section II.B.1.

In developed countries, conservation organizations generally do not have to deal with indigenous communities who are living and surviving off the natural resources in areas designated for conservation. However, this does not mean that the problems identified in developing countries are not also found in the developed countries, rather they have not driven the approaches being used in conservation. In developed countries, greater emphasis has been placed on developing the tools to protect and restore habitats and species, and in eradicating invasive species that are threatening indigenous species composition and altering succession. The spread of invasive species and the concomitant loss of indigenous species is a problem driven by international trade and the globalization of the world. The changing transportation networks and the development of a global economy are contributing to an accelerated rate of the spread of invasive species that has not been documented prior to this time. In addition, the spread of invasive species frequently occurs in ecosystems that have been significantly modified by human land use activities and few good control mechanisms have been produced. According to P. M. Vitousek, humans have been modifying their environments at unprecedented rates during the last several decades. Because of the role that humans are playing in spreading species and in altering the health of ecosystems, one of the greatest challenges to the successful implementation of conservation projects may be our ability to control the spread of invasive species. The impact of invasive species on conservation efforts in developed countries is discussed in Section II.B.2.

Some of the direct and indirect impacts of humans are much more difficult to manage in conservation since we are just beginning to understand these effects. Many of these effects occur at longer temporal and spatial scales, making it more difficult to explicitly link the cause-and-effect relationships so that tools could be identified to manage ecosystems and species in ecosystems. Few good solutions have been produced for even assessing and managing these constraints, and so useful tools are not available to transfer to conservation at this time. The impacts of humans on species loss and changing ecosystem structure and function will be further discussed in the following sections.

1. Humans and Sustainable Development as an Integral Part of the Conservation Formula

Within the last two decades, there has been an increasing understanding that conservation projects cannot be successful when implemented in isolation from neighboring human populations. Maintaining nature reserves has undergone a major strategy shift from trying to keep people out of reserves to allowing some economic development by local communities as an integral part of conservation strategies (discussed in the works of K. Brandon, F. H. Buttel, R. J. A. Goodland, J. McNeely, M. L. Shaffer, M. Wells, and P. C. West). The response of the international community in dealing with this issue has been to design projects that explicitly link international aid to local communities with the establishment of nature reserves to conserve species. The goal of these projects has been to improve the livelihood of people living close to the natural resources so that they would not be dependent on extracting species or resources from the protected area. The strategy behind this approach was to find a balance between development and conservation by providing alternative approaches for people to survive and at the same time provide zones of lower human impact for conservation purposes.

This new model of conservation attempts to reconcile human pressures on resources in areas that are rich in biodiversity and therefore have been identified as high priority areas for conservation. This represents a major shift from the traditional conservation model, whose main focus was to lock up resources by eliminating human presence in nature reserves without considering its impact on the survival of the local communities excluded from these areas. Conservationists now recognize that successful application of their models will require explicit recognition and alleviation of human dependence on natural resources at the local level. It has become quite apparent that it was insufficient to set up borders and assume that people with few options would stay out of these areas.

In fact, humans have had an integral hand in shaping the very land that we now wish to conserve in many tropical countries. It is important to recognize that present forest conditions may be a relict of past management and land use activities of the indigenous and rural communities, so that eliminating them from the landscape may actually change the forest composition to one that may not be acceptable or desirable for conservation purposes. In wet tropical forests, humans have strongly influenced the resident biodiversity; indeed, the "virgin forests" and exceptionally rich biodiversity that are the focus of conservation efforts in the Amazon River basin are in part relicts of old agricultural practices and management (see W. Baleé, M. Pinedo-Vasquez, D. A. Posey, and H. Raffles).

The importance of the human dimension was especially reinforced for conservationists working in the tropics, where desperate people continued to harvest resources from forests set aside as nature reserves for conservation purposes (see works of K. Brandon, M. Dove, J. McNeely, and M. Wells). Many of the conservation organizations funding the establishment of protected areas quickly recognized that nature reserves could not be maintained if they were designed to exclude local communities without providing alternative sources of income. Most of the current projects designed to conserve biodiversity are encouraging local communities to use areas outside of reserves or at least along the buffer zones of reserves. This strategy resembles the UNESCO Man and the Biosphere (MAB) design for reserves, which is based on having concentric zones representing different intensities of human use—with the center or core area designated for conservation without human presence. However, shifting or confining people's use of a forest to particular zones without their cooperation is impossible in most developing countries. Often, land tenure or ownership does not exist for these local communities. Furthermore, these communities are accustomed to adjusting their resource acquisition efforts within the landscape both seasonally and by migrating to different locations.

As part of the new conservation approach, ecotourism and the harvesting of nontimber forest products have often been presented as two viable alternatives for local communities to generate income (see J. L. O'Hara). Yet so far these options have not proven to be effective in meeting the dual goals of biodiversity conservation and rural economic development (see M. Kremen, A. M. Merenlender, and D. D. Murphy).

Thus social factors can play a major role in determining the effectiveness of reserves, that is, the pressures of human activity and population growth will probably be the major determinants of whether a reserve will successfully conserve a given species (see A. Dobson). Sometimes the scientific knowledge of how much area is needed by a species cannot be implemented in the face of these social pressures. In these cases, there is nothing wrong with using the science to determine the necessary size of the conservation reserve, but managers must keep in mind that social and economic variables are exerting more control on whether the reserve will be able to protect its resident biodiversity.

2. Humans and Invasive Species

Within the last decade, many conservation organizations have implemented programs to control or eradicate invasive species, especially those that are exotic. This new objective is driven by the many field examples of invasive species disrupting or altering native ecosystem functions and structures, as well as their biodiversity. The significant economic costs of eradicating, or attempting to eradicate, invasive species have highlighted the difficulty of managing such species. One of the biggest challenges of future conservation efforts will be the ability to control or minimize the presence of invasive species, because they will hamper or limit the options that are available for the conservation of indigenous species and their intact ecosystems. The spread of invasive species will probably become more critical in the future as humans create the conditions or change disturbance cycles sufficiently to give invasive species the opportunity to dominate invaded habitats.

Humans have played a dominant role in the spread of invasive species around the world. Humans function as vectors of invasive species and/or alter the landscape in a way that enables invasive species to become better competitors than indigenous species. Many weeds, insects, and other animal pests are transported accidentally with produce and goods; food crops and horticultural varieties are often introduced to new environments intentionally for agriculture and landscaping. Likewise, there is a long history of humans introducing game animals to new territories. Once invasive species are introduced into a landscape, many of our land use activities have been implicated in facilitating their spread. There are several possible mechanisms by which the rate of invasive species spread is enhanced: (1) the combined effect of land use change and an invasive species may alter disturbance cycles to which indigenous species are adapted and (2) invasive species disrupt established abiotic resource availabilities within an ecosystem that results in the ecosystem conditions becoming more favorable for exotic species (see Office of Technology Assessment, 1993; as well as the work of J. Martinelli, J. Randal, P. M. Vitousek, M. H. Williamson, and others).

Although the invasion of ecosystems by alien species certainly has occurred throughout human history, the problems of alien invasion remain particularly challenging both for scientists studying patterns and processes of invasion and for ecosystem managers attempting to eradicate or control invaders. In many parts of the world, particularly on islands, even the greatest conservation efforts are ineffective at countering the proliferation of exotic species. Where invaders are well established, such as rats in New Zealand and melastome plants in Hawaii, invasive species are nearly impossible to control and conservation strategies need to explicitly manage invasive species that currently cannot be eliminated. In less severe cases, conservation practices may be able to protect biodiversity and limit the establishment and spread of exotic species.

What is needed to adequately handle invasive species and prevent their establishment is a framework for understanding what qualities predispose an ecosystem to invasion and what species characteristics make invaders so ecologically opportunistic. Such a framework must be developed from knowledge of rates and types of species invasions, establishment conditions, and mechanisms of species dispersal in new habitats (see work by M. J. Crawley, and others). Clearly, the inherent dispersal ability of a species is a key determinant of its invasion potential. With respect to dispersal, classic ecological theory identifies two general types of species: those that occupy stable habitats at relatively constant densities and those that are locally ephemeral and have extreme population dynamics. Naturally, the selection pressure for dispersal ability among this latter group is high, and species of this type are well represented among introduced species. Of the plants in this group, many are early-successional species such as grasses and legumes. Probably more important than species', natural dispersal mechanisms, however, are their associations with humans. For example, stocks of seed and produce shipped overseas often carry with them seeds of many weed species that germinate in their new environment. These shipments commonly include alien microbes and animal pests as well. Intentional introductions of species imported for agriculture and other human uses are common, as are introductions of organisms employed as biological control agents to stanch the tide of previously introduced and invasive species.

Whether or not an introduced species establishes itself in a new ecosystem depends on its ability to meet a specific invasion criterion, or population growth rate, upon introduction. With an increasing body of formal ecological theory, calculations of invasion criteria may be more or less spatially and temporally explicit, including information about generational overlap and demographic and environmental stochasticity. Godfray and Crawley have provided a thorough review of invasion criteria, including temporal and spatial complexity. However, because of the complexity of invasion criterion models, it is often difficult to predict whether or not a specific introduced species will establish itself without actually conducting an invasion. The risk of such an experiment is understandably high, so in general it may be wise to assume that any introduction is irreversible.

Although some introduced plants and animals commonly fare poorly in new habitats, there are many examples of introduced species that better fit their invasion criteria in their new range than in their original environment. Frequently this increase in fitness in an invasive species follows the release from pressures of herbivory, predation, or a pathogen associated with the invader's native environments. Alternately, an introduced species may fail to invade if it is a relatively poor competitor in its new environment or if it is preferentially grazed or predated compared to native species. An oft-cited example of success following herbivory release is that of the invasion of New Zealand by Monterey pine (*Pinus radiata*). This pine is a native of the Monterey peninsula of California, where it tends to grow as a small, scrubby tree. When imported to New Zealand as an experimental timber species earlier this century, it was discovered that the pine grew many times taller and much straighter, and was longer-lived than in its home range (see Ledger and Richardson). With the help of widespread planting, seedlings of *P. radiata* and other American pines and firs quickly escaped from where they had been originally planted. Today the spread of these wildlings is argued to be the dominant ecological threat to many alpine and lowland forest communities in New Zealand.

In addition to a species' potential for invasion, conservation scientists and practitioners must also consider the predisposition of a community to be invaded. Research by Crawley and others found that recently disturbed and less contiguous communities are more readily invaded than closed and late-successional communities. It is very significant that these characteristics are associated with areas of high human activity. Species already associated with areas of human development are consequently more likely to be introduced to new areas by human vectors, and the areas to which they are introduced tend to be more readily overwhelmed. Observations suggest that alien plant species richness correlates positively with proximity to human development and transport centers.

Once an invasive species has successfully colonized a new habitat, the ecological consequences of the invasion are of greatest concern to conservationists. Again, although prediction of the exact consequences of a particular invasion may be difficult, a large body of current ecological theory may be harnessed to make some generalizations. From a community ecology standpoint, the introduction of a new species could have a large number of direct and indirect effects on other populations depending on the strength of its competitive interactions and its trophic position. For example, an introduced plant that successfully competes for a limiting resource with another plant might cause population declines of a herbivore that grazes exclusively on the native plant. Alternately, if an invader does not necessarily exclude a competitor but is palatable to a herbivore that also grazes on the competitor, the herbivore population would be expected to increase and the population of the competitor would be expected to decrease, mostly as an indirect consequence of the invader. Of course, when an invader takes a new, high trophic position in a community (e.g., rats, mustelids, and cats on South Pacific islands), it is certain to have a dramatic effect on prey that have not evolved under the selective pressure of predation.

It is also important for conservationists to consider the consequences of invasive species on patterns of succession. As noted earlier, open and disturbed communities are commonly prone to invasion. If an invasive species or suite of species is characterized by an early-successional stage in its native range, its invasion following a disturbance may effectively arrest succession on the site, both excluding native species in the present sere and precluding species typical of later stages of succession. Several examples of this have been recorded in Hawaii with an invasive nitrogen-fixing plant and for warm-season grasses that are arresting succession (see works of P. Kareiva, P. M. Vitousek, P. Walker, and others).

Invasive species are also known to have dramatic effects on the abiotic environment of the newly invaded ecosystem. Forest invasions by perennial grasses have been shown to increase the frequency and intensity of fire disturbances and thus shift the annual timing of these events. Invasive trees on forested watersheds have been known to transpire much more water than native trees, significantly altering forest hydrology (see R. M. Cowling).

Of ultimate importance is the degree to which the successful invasion of an organism affects directly and indirectly the community as a whole and the ecosystem functions associated with it. In many ecosystems, ecologists have observed that certain species, known as keystone species, play a critical role in shaping overall community structure. The exclusion of keystone species often results in a cascade of population shifts of directly and indirectly related species. R. Paine and others have illustrated this in a marine intertidal community by removing a top predator and noting a general decrease in total species diversity due to the overwhelming growth of a specific prey population. Considering such observations, conservationists must be particularly sensitive to the possibility that an invasive species may outcompete, predate, or otherwise exclude a keystone species, resulting in a much larger loss of biodiversity.

Because of the uncertainty in our understanding of invasive species, conservation biologists will have to consider whether invasive species are present in areas being managed for species and their habitats. If invasive species are present, it will be important to consult the biological literature and determine whether and what types of effects may be attributed to that invasive. If these invasive species are shown to negatively affect indigenous species and ecosystems, it will be critical to eradicate small pockets of invasive species when they are found to prevent their establishment and spread to the point where they cannot be controlled. This suggests that conservation practices will have to aggressively manage invasive species even when their impacts have not been conclusively shown to detrimentally affect the target system. It will be critical for conservation to establish priorities and develop an effective strategy for preventing and handling invasive species. Several ecosystems in which conservation projects had to consider eradicating or controlling invasive species are given in Table II.

III. TOOLS FOR MANAGING AND CONSERVING SPECIES

When Michael Soulé, the "father of conservation biology," was interviewed in 1994 and asked about the developments in conservation biology in the last decade, he suggested that there has not been a revolution in how conservation biology is practiced. Rather, he stated that conservation has been experiencing a gradual refinement in the principles and tools of the field. Soulé suggested that the major tools have been the application of theory from the fields of island biogeography, population biology, and community ecology, namely, consideration of the matrix landscape within which a habitat is embedded, the development of biological corridors, the development of new tools in molecular biology, and the development of integrated population viability analysis frameworks that include demographic, environmental, and genetic stochasticity. All of these are commonly used in conservation today.

Since the early 1990s, however, there has been a revolution in the philosophical basis of conservation biology. As discussed earlier, five philosophical changes have occurred in conservation, however, two have arisen to the forefront of discussions within the last decade. These two philosophical changes have also been the most difficult to convert into practical applications for natural resource management. One was the acceptance of ecosystem management by United States federal agencies to help resolve conflicts arising from public perceptions of poor management of natural resources. The second philosophical change was the acknowledgment of the importance of human dimensions in natural resource management. The tools for implementing these holistic approaches to ecosystem analysis are rapidly developing so that they can be practically applied in the field. Because these holistic approaches to species conservation are still evolving, many ecologists are not comfortable with them (see D. Wilcove). Despite these reservations, many conservation organizations are actively developing the ecosystem approach as their primary tool for managing species. By accepting these new approaches, conservation biology will be able to contribute to the development of application tools for studying systems holistically and for integrating the social and natural sciences. Unfortunately, conservation biologists will not be able to borrow well-tested tools from these other disciplines since consensus is still evolving on which suite of tools best allows theories to be converted into practical applications.

Approaches to conservation continue to evolve as new knowledge is produced in population, community, ecosystem, and landscape ecology, and in social ecology. Some of the change in conservation applications occurred from a greater understanding of several of the reasons why the Endangered Species Act has had limited success in halting species extinction or contributing to their recovery. The Endangered Species

TABLE II

Examples of Areas Where the Conservation of Natural Habitats and Species Is Threatened by Invasive Species

Conservation target	Invasive threat to conservation efforts
Wet lowland and alpine o'hia (*Metrosideros polymorpha*) forests, Hawaii, USA	Among numerous invasive species, the Latin American native plant *Miconia calvescens* grows rapidly to dominance over native stands where established. The "green cancer," as it is known in Tahiti, reproduces prolifically with high success contributing to its rapid dispersal. Miconia threatens biodiversity at all levels in these forests because of the high degree of endemism typical of the island ecosystems.
Fraser fir (*Abies fraseri*) forests, Great Smoky Mountains National Park, Tennessee, N. Carolina, and Virginia, USA	The European balsam wooly adelgid (*Adelges piceae*) is a sucking insect that kills firs over several years by draining their sap. Already the invader has eliminated more than three-quarters of all spruce-fir forests in the southeast US, shifting dramatically microclimate, vegetation, and food-web dynamics. A number of plant and animal species have been listed on the federal endangered species list as a direct consequence of habitat destruction in the national park.
Wetland plant communities, Everglades National Park, Florida, USA	The Australian *Melaleuca quinquenervia* forms dense monotypic stands in marshlands, reducing native plant diversity by 60 to 80%. Presently it covers more than 450,000 acres, and its rate of spread has grown to 50 acres per day. It is especially problematic to conservation practitioners because stress and damage from fire and herbicide cause it to release abundant seeds.
Freshwater bivalve communities, Mississippi River basin, USA	The zebra mussel (*Dresissena polymorpha*), a recent invader from the Caspian and Black Seas, adheres to the shells of freshwater mussels—as many as 10,000 zebras to one native—and interferes with the feeding, growth, movement, respiration, and reproduction of these species. The Mississippi basin hosts more native mussels than any other river system in the world, a community largely responsible for nutrient cycling and sediment mixing in this system. Zebra mussels have no American parasites or predators.
Shortgrass prairies and rangelands, Montana, Wyoming, the Dakotas, and Nebraska, USA	Leafy spurge (*Euphorbia esula*), a Eurasian herb, is a grassland invader known to completely and quickly displace prairie grasses and forbs by overshading neighbors and outcompeting them for available moisture and nutrients. It is also known to reduce productivity of western grazing lands by 50 to 75%. The spurge is difficult to eradicate owing to its deep root systems, high seed production, and dispersal abilities.
Tropical forest birds, Guam (also Hawaiian and other tropical and subtropical communities).	The brown tree snake (*Boiga irregularis*), a native of several South Pacific islands and northern Australia, is a widely adaptable predator that exacts its heaviest toll on tree-nesting bird species. As dense as 12,000 individuals per square mile in some forest areas, the brown tree snake has already eliminated 9 of Guam's 11 native forest birds and most of its exotic birds. Although not yet naturalized in Hawaiian forests or other likely systems, the nocturnal snake is noted for stealth and has been found on several occasions in cargo transported by air to Hawaii and North America.
Riparian woodlands, Sonoran Desert and elsewhere in southwest USA	The Eurasian tamarisk tree (*Tamarix spp.*) is the superior competitor for water in these fragile desert ecosystems and devastates soil structure and hydrology, thereby altering the habitat of numerous plants and rare birds, mammals, amphibians, and invertebrates. It has no native predators or diseases, and the selective pressures of widespread grazing on native cottonwoods and willows exacerbate its proliferation.

Act has tended to protect high-profile species rather than overall biodiversity and did not sufficiently protect habitat reserves for recovered species. The focus on species conservation as the primary tool has been shown to inadequately reflect whether a habitat was sustainable. Conservationists moved to an ecosystem approach when it became obvious that species by themselves were generally poor indicators of ecosystem function and resilience. A brief discussion of how the philosophical changes in conservation have been translated to field applications will highlight the shift occurring in conservation.

A. Models

Past modeling efforts focused on using data-intensive species population and community models, such as Minimum Viable Population (MVP) and Population Viability Assessment (PVA) models (see S. Beissinger, M. S. Westphal, and others), to understand how to conserve species. These models focused on the single species without taking into account the ecological factors of the species' habitat. As discussed earlier, these types of approaches have not been adequate to predict the ability for conserving single species in ecosystems and landscapes. Recently, PVA models have been adapted to consider the spatial context, to better incorporate uncertainty and stochasticity, and to have more statistical rigor. There is also a growing attempt to develop wildlife habitat association models and link these with GIS databases to predict species diversity in actual geographic locations.

Models are also evolving as tools to help balance trade-offs between the design of conservation areas and conflicting natural resource uses. New models are using multiobjective programming to examine the multiple trade-offs that can exist in the design of a reserve (see K. D. Rothley, O. J. Schmitz, and others). To design bioreserve networks, these models are incorporating decisions that are made by animals for food and shelter, by people interested in recreation, and also by landowners wanting current income.

It is unlikely that any reserve or reserve network can satisfy the demands of all of the stakeholders involved with the land base within and/or surrounding reserves. So planners must find ways of reaching favorable compromises among competing demands. In many cases, however, reaching compromise solutions may require considering a multitude of options. For example, if a goal of reserve network design is to select 5 reserve areas from a set of 20 potential sites on a given land base, then one must decide among 15,504 possible networks of 5 reserve areas. The decision-making becomes even more complex when one must consider additional criteria such as size and shape of reserves, interconnectedness of reserves and the need for corridors, numbers of species preserved, locations of the reserves, and the cost of purchasing land (as discussed by G. Caughley, A. Gunn, and K. D. Rothely). Reserve planning has become increasingly sophisticated and it now enlists numerous algorithmic and optimization tools that help reconcile trade-offs among competing alternatives.

B. Species and Their Spatial Relationships in the Landscape

Much of the early theory development by conservation biologists was based on determining the relationships between species and their space (as first articulated by R. H. MacArthur and E. O. Wilson). The suggestion that habitat area was positively correlated to species richness provided arguments for designing larger natural reserve areas for maximizing the conservation of species. The size of a reserve area needed for effective conservation has been argued in the literature for more than 20 years without the development of a consensus. Part of the reason for a lack of consensus is that we know that area by itself is insufficient to characterize how many species can be maintained at a site (see work by J. F. Franklin and D. A. Saunders for discussion of spatial scale issues).

Even though space has been an important theoretical issue in conservation, the practical application of theory in conservation appears to have no dominant spatial scale of analysis. Therefore, theories and applications have not been linked at the spatial scale. A lack of defined spatial scale in applied conservation has been strongly influenced by the focus on specific animals as important species to conserve. The spatial scale at which conservation has been practiced has varied according to the territorial needs of the species of interest. This trend is apparent from a survey of key journals published in 1998 that tabulated the scales of analysis used by conservation biologists, community and population ecologists, and social ecologists. An amazingly high proportion of the articles published in the journal *Conservation Biology* did not even present the scale being used in the study; this suggests that spatial scale was not considered relevant by the authors for understanding their system. When a scale of analysis was presented, the smallest scale (<-0.01 ha) was as equally represented as the largest scale ($>-10{,}000$ ha). Again, this reflects the focus on animals.

This contrasts with much of the information collected from community and population ecology journals, where small-scale research is predominantly used to understand species relationships to ecosystem parameters. For example, plant ecologists who are developing the theory behind species dynamics focused on scales that were less than 0.01 ha in size. Some studies are being conducted at larger scales, but these larger scales of analyses are applied in the tropics and not other regions of the world. In the humid tropics, the presence of rare plant species has necessitated the establishment of permanent plots that are 50 ha in size (see P. Ashton, S. Hubbel, and others).

If conservation biologists are going to accept the ecosystem management paradigm, they will not be integrating the scales of analyses that are particular to ecosystem ecology. The watershed has been accepted as the relevant scale in ecosystem management (see FEMAT, 1993), but this is only one of the many scales that could be used in conservation. Selecting one scale of analysis for all systems is a real problem because scale is site dependent and must be chosen to reflect the driving variables controlling the functioning and resilience of that system.

C. Reserve Designs

Reserve design used to be based on defining the carrying capacity of an appropriate area needed by a species to maintain it in the landscape. This type of approach to defining the area was sufficient when there was an immediate need to set aside areas to protect species. However, this type of approach does not deal with the fact that the quality of the space makes a difference in how effectively it is used by the species—it is not just a question of total area and the amount of interior and edge environments. We realize that the distinction between interior and edge environments appeared to be easy to distinguish in the field, but in fact was not so clear when species other than birds were the species of conservation interest. How the quality of an environment changes with respect to the matrix landscape that is not in reserves is receiving considerable research attention today.

It is now universally acknowledged that the conservation of species requires the preservation of natural habitats within ecosystems. One fundamental concern is that if the rates of habitat loss due to resource exploitation continue at current levels, many habitats will disappear forever by the middle of the twenty-first century (see D. S. Hik, O. J. Schmitz, A. R. E. Sinclair, and others). The growing concern over the loss of habitat, and the concomitant loss of biodiversity, has generated intense interest in the use of reserves and reserve networks as conservation tools. For example, the Wildlife Conservation Society's flagship project, "Paseo Pantera," is an effort to consolidate and link protected areas throughout Central America. The intention is that these reserves will afford species an extensive refuge from exploitation.

The most fundamental strategy in reserve design is to circumscribe and set aside parcels of land that are key sources of biotic diversity, and to exclude all human activity within their boundaries. In this sense, reserves act as repositories for much of the world's biota in perpetuity. There are two main reasons why such a strategy alone is insufficient. First, natural habitats decline both within and outside of reserves through natural processes (e.g., succession) and as part of the differential impact of disturbances within the matrix landscape. Thus, habitat within protected areas will continue to change, albeit at slower rates than outside of protected areas. Consequently, the structure and function of those putatively protected ecosystems will not be sustained over the long term simply by delineating them as reserves (as discussed by A. R. E. Sinclair and others). Second, setting aside land to protect specific ecosystem types may no longer be possible because the land base has been degraded to a point where the complement of habitats needed for a sustainable ecosystem no longer exist. The solution is to adopt a new conservation strategy that promotes the reestablishment of habitat via habitat renewal and restoration at rates that offset natural habitat decline within a landscape framework. Accordingly, management must choose not only the size of land parcels to preserve, but also the size of area that balances habitat loss with habitat renewal. This may require the restoration of some ecosystem types and the eradication of invasive species.

Thus reserve designers must adopt a holistic perspective that considers habitat decline and renewal as consequences of succession and disturbances. They must also maintain a long-term, large-scale perspective in deciding how to balance renewal against loss. For example, consider a 100-ha parcel of old-growth forest that forms part of a reserve that is rapidly decreasing in area due to natural stand break-up (see B. C. Larson and C. Oliver). In order for old-growth forest to be sustained within the reserve, another 100 ha of younger successional forest must be available to replace the declining stand as that older forest stand is shrinking, not 100 or 200 years in the future. In this case, one must set aside much more land than 100 ha and protect a wider variety of successional-age stands in order to have a sustainable old-growth forest ecosystem. In addition to these biologically motivated criteria for reserve design, there is growing recognition that human social factors outside of the reserves can play a major role in determining the conservation effectiveness of a reserve.

D. Economic Evaluations

In the past, economics was the dominant tool being used to build a market-based rationale for conserving species and ecosystems. Economics was used to show that conservation, and the resulting protection of ecosystem services, is more valuable than extraction or harvesting of resources. However, this overreliance on economics as the primary tool to build a rationale for conservation has produced mixed results (see C. H. Freese and others).

In the late 1980's, economists provided conservation biologists with what appeared to be an excellent tool to conserve tropical forests that were high conservation priorities. The analyses also provided different options for generating income by producing values for previously non-market-valued products. Historically, tropical forests had been undervalued when only timber and agricultural uses were used to estimate their net worth, so that there was no clear economic incentive to conserve these forests (see P. S. Ashton, T. Panayotou, C. M. Peters, and others). These authors have suggested that the lack of proper economic valuation has contributed to accelerating the rate of harvesting and conversion of forests for other uses. Therefore, a tactic to decrease deforestation rates was to determine the "real" value of the forest by including all other products that could be extracted from it (e.g., nontimber forest products and services like watershed protection). Using this analysis, the forest was no longer undervalued, and alternative products could be extracted that would provide income to local communities. This strategy is now widely promoted by the international development community (see K. Brandon, R. J. A. Goodland, and M. Wells). Its goal is to conserve forests and biodiversity in reserves while also generating income for local communities from the same reserve or surrounding areas.

This strategy of combining conservation with sustainable development was supported by studies reporting that extraction of nontimber forest products would return greater value in the marketplace than timber harvesting. In addition, this strategy was promoted as a new alternative to reduce the dependence of the local community on natural resources that were being protected in a newly formed conservation area. International funding organizations were interested in this strategy because it suggested that the goals of conservation were compatible with developing an economic base for local communities, who could forgo harvesting resources (trees, wildlife, etc.) within designated reserve areas if sufficient income could be generated from extracting other resources or from providing recreational opportunities from the forest. Unfortunately, it is not clear whether the goals of conservation and sustainable development are compatible. M. Dove has conducted research suggesting that this strategy should be approached with caution, since developing a total value for a forest does not protect it from being overharvested.

Some economic analyses have been criticized for producing unrealistic expectations of the value of what can be sustainably extracted from a forest (see Bawa, Godoy, and others). Biological inventory data do not indicate the amount of a product that can be harvested in an ecologically sound manner, and also significantly overestimate the expected income generated annually from extracting these products (Belsky, Pinedo-Vasquez, and Siebert). Indeed, placing an economic value on a resource can increase the demand for that product so that it is quickly overharvested; when this occurs, the species may be lost from the landscape where it previously was quite common (see J. L. O'Hara and M. Pinedo-Vasquez).

IV. FUTURE DIRECTIONS AND CHALLENGES IN CONSERVATION

The solutions to the problems facing conservation biologists will not be found solely by adopting new paradigms such as ecosystem management or sustainable development. Blending the theories developed in conservation biology with these new paradigms will potentially allow for better management of species within their landscapes. However, conservation biologists will not automatically find the tools that they need from the different disciplines. These new paradigms are currently struggling to develop tools that will be able to take theories and make them applicable to, for example, forest certification. Furthermore, conservation biologists have not developed structures similar to ecosystem management that would allow them to develop an adaptive management model (i.e., a way to allow scientific data to continuously modify policy links as better information became available). Without such a process or framework to effectively link policy to science and management, it is difficult to integrate scientific findings into management itself. As stated by J. C. Gordon in 1999, "Often, policy is made with 'old' science, while interest groups confront the manager of public resources with 'new' science." One could speculate that "old" science is used because the tools have not been developed to implement or make the "new" science work at the ground level or in policy.

Conservation biologists will also have to confront the fact that humans have altered the environment on a global scale (see NRC, 1992). There is probably no place in the world that has not experienced human impact. In many cases, the current system that we study and want to conserve is something that may be an artifact of past human activities. If this is not recognized, it may be extremely difficult to conserve species because we really do not understand the driving variables that may control the resistance and resilience characteristics of the system. As protected area managers are increasingly called upon to generate funds through development projects such as ecotourism, timber harvest, and nontimber forest product extraction, it is important that the original goal of setting up these reserves as "genetic libraries" is not compromised.

Restoration is frequently suggested as a way to reestablish desired ecosystems and at the same time maintain the biodiversity of indigenous species and exclude invasive species. This is especially relevant considering the problems that conservation managers are facing with invasive species and altered landscapes. However, the science of restoration ecology is still evolving, and standard tools are not readily available for conservation purposes. A current practice has been to allow the development and elimination of a specific habitat with the assumption that an equivalent habitat will be recreated at another location. This approach has been commonly used in wetlands. Unfortunately, the successful restoration of a habitat that can then be used by an endangered species rarely results. For example, several authors (J. B. Zedler and R. Langis) found that a restored marsh that was supposed to support an endangered species was not suitable habitat for that species. They also found that after five years the marsh was still not supporting other species found in their reference marsh. Restoration ecology is still a young science, and it is not surprising that restoration tools have not had sufficient time to be tested in a variety of systems. Many restoration efforts are currently dealing with eradicating invasive species that have begun to dominate much of our landscape. Yet we have a minimal understanding of the impact of invasive species within landscapes and of the long-term ecological effects that they may exert in these newly modified environments.

A central challenge for conservation in the twenty-first century is to recognize that conservation efforts will be ineffective if the science in conservation biology fails to incorporate the human factor. Most models linking the human dimension to conservation are theoretical and few have managed to realistically link people and the environment. At this time, the tools are not readily available to link applications of the social and natural sciences, and it will require some innovative approaches to address this problem. Several other problems must also be explicitly addressed in conservation:

- New paradigms that can help shift conservation from a species focus to a holistic approach have good theoretical grounding, but there are few frameworks for implementation.
- Past land uses and their legacies have already modified the environment (i.e., habitat fragmentation, chemical pollution, and introduction of species that have become invasive) sufficiently that ecological theory or conservation models that ignore the prevalent effects of humans on the environment will not succeed in field application.

Conservation biology has done well in developing the science of understanding individual species in their habitat, performing spatial scale analyses of individuals, and modeling their activity within the landscape. It has not done well in the following areas:

- Incorporating the roles and needs of humans in conservation situations, especially at the local level. The values and principles of national and international organizations have largely driven conservation efforts.
- Developing frameworks for moving beyond the species focus and incorporating the dynamic nature of ecosystems into these frameworks.
- Focusing on the use of "easy" tools (i.e., GIS models) has restricted the diversity of conservation approaches and created a dependence on a small set of tools, thereby excluding the use of a more varied and innovative suite of tools.

See Also the Following Articles

BIODIVERSITY-RICH COUNTRIES • CONSERVATION MOVEMENT, HISTORICAL • ECONOMIC VALUE OF BIODIVERSITY, OVERVIEW • ECOSYSTEM, CONCEPT OF • HOTSPOTS • INTRODUCED SPECIES • KEYSTONE

SPECIES • LAND-USE ISSUES • RESOURCE PARTITIONING • SUSTAINABILITY, CONCEPT AND PRACTICE OF

Bibliography

Caughley, G., and A. Gunn (eds.). (1995). *Conservation Biology in Theory and Practice*. Blackwell Science, Cambridge, Massachusetts.

Daily, G. C. (ed.). (1997). *Nature's Services. Societal Dependence on Natural Ecosystems*. Island Press, Washington, D.C.

FEMAT (Forest Ecosystem Management Assessment Team). (1993). Forest Ecosystem Management: An Ecological, Economic and Social Assessment. U.S. Govt. Printing Office no. 1993–793–071. Washington, D.C.

Hunter, M. L., Jr., (ed.). (1999). *Maintaining Biodiversity in Forest Ecosystems*. Cambridge University Press, Cambridge, United Kingdom.

Jianguo, L., and W. M. Taylor (ed.). (2000). *Integrating Landscape Ecology into Natural Resource Management*. Cambridge University Press, Cambridge, United Kingdom.

NRC (1992). Global Environmental Change: Understanding the Human Dimensions. (P. C. Stern, O. R. Young, and D. Druckman, eds.). National Academy Press. Washington, D.C.

OTA (Office of Technology Assessment). (1993). Harmful Non-indigenous Species in the United States. U.S. Govt. Printing Office. Washington, D.C.

Primack, R. B., D. Bray, H. A. Galletti, and I. Ponciano (eds.). (1998). *Timber, Tourists, and Temples: Conservation and Development in the Maya Forest of Belize, Guatemala, and Mexico*. Island Press, Washington, D.C.

Sutherland, W. J. (eds.). (1998). *Conservation Science and Action*. Blackwell Science, Oxford, United Kingdom.

Vogt, K. A., J. Gordon, J. Wargo, D. Vogt, H. Asbjornsen, P. A. Palmiotto, H. Clark, J. O'Hara, W. S. Keeton, T. Patel-Weynand, and E. Witten. (1997). *Ecosystems: Balancing Science with Management*. Springer-Verlag, New York.

Vogt, K. A., B. C. Larson, D. J. Vogt, J. C. Gordon, and A. Fanzeres (eds.). (1999). *Forest Certification: Roots, Issues, Challenges and Benefits*. CRC Press, Boca Raton, Florida.

CONSERVATION MOVEMENT, HISTORICAL

Curt Meine
International Crane Foundation

GLOSSARY

anthropocentric Human-centered in perspective, especially with regard to the value of the natural world.

biocentric Valuing the existence and diversity of all biological species.

conservation biology An integrative approach to the protection and management of biological diversity that uses appropriate principles and experiences from the natural sciences, the social sciences, and various resource management fields.

preservation The school of conservation philosophy that emphasizes the protection of natural features and landscapes from human exploitation.

Progressive Era The period in American political history, especially coinciding with the presidency of Theodore Roosevelt (1901–1909), in which the conservation movement gained definition.

sustainable In general, capable of meeting economic goals in a manner that does not degrade the quality of the underlying ecosystem

sustained yield The management and harvesting of renewable resources in a manner that does not exceed their rate of replacement or reproduction.

utilitarian The school of conservation philosophy that emphasizes the value of the natural world in contributing to human well-being, especially in terms of economic standard of living.

MODERN EFFORTS TO CONSERVE BIODIVERSITY have their foundations in older traditions of resource management and nature protection. This chapter traces the history of the conservation movement, focusing on those events and patterns that led to the emergence of biodiversity conservation from earlier utilitarian and preservation-oriented approaches. Because the conservation movement continues to redefine itself, this article concludes with a consideration of key themes from recent history.

I. THE BACKGROUND OF CONSERVATION

As a conscious goal of citizen action, public policy, and professional endeavor, conservation first gained coherence in the late 19th and early twentieth centuries. Over the decades, the conservation movement has evolved in response to varied forces: emerging findings in the natural sciences, new environmental threats, shifts in philosophical assumptions and aesthetic standards, novel technologies, expanding legal mandates, and changing social, economic, and political conditions.

The complex interplay of these forces has produced a movement that remains in flux and whose goals continue to evolve, even as the world's ecosystems face increasing challenges to their integrity.

A. Prehistoric Precedents

The term *conservation* acquired its modern meaning in the early twentieth century, especially with the rise of the Progressive Era conservation crusade in the United States. However, as an expression of cultural commitment to an enduring relationship with the natural world, conservation has much deeper historic and prehistoric roots.

Prehistoric peoples did not live in a simple or constant state of peaceful coexistence with and within their natural surroundings. Converging lines of evidence from paleontology, paleoecology, archaeology, and anthropology suggest in fact a sobering picture of the human past: that the dispersal of the human population over the past 120,000 years has been accompanied by spasms of continental and insular extinction and other forms of environmental degradation. The best known of these prehistoric extinction events involves the disappearance of mastodons, ground sloths, camels, giant beaver, and other large North American mammal species coincident with a major wave of human colonization between 12,500 and 10,000 years ago. The same general pattern of extinction has been traced back over the centuries as humans moved out of Africa and Eurasia into Australia, Polynesia, and island groups around the world. The implications of this pattern of "dreadful syncopation" are controversial, and research continues on the exact sequence of cause and effect, the interaction of probable causal factors, and the precise mechanisms of extinction (MacPhee, 1999).

In contrast to this record of prehistoric anthropogenic extinction, however, there is countervailing evidence from the many cultures that have achieved relatively sustainable ways of life. Native peoples in landscapes throughout the world developed sophisticated belief systems and land-use traditions that recognized their connections to and dependence on the nonhuman world. In the past, these traditions allowed people to adapt to even extreme environments throughout the world. The Maasai of East Africa, for example, practiced livestock herding practices in a manner that allowed them to coexist with the native megafauna of their semiarid homeland. Arctic dwellers coexisting with caribou herds and marine mammals, shifting cultivators and agroforesters in the humid tropics, small-scale farmers in temperate zones, fishing villages in coastal, riverine, and wetland environments: these and other traditional cultures developed social systems that balanced human needs against the capacity of the natural environment to meet those needs. Some of these traditions have survived into the modern era, though they must contend with increasing pressures from diminished resources, population growth, altered land tenure systems, and rapid economic and technological change.

The resource-use practices of native peoples had significant impacts on the biota, though these impacts varied in intensity over time and according to ecosystem type and taxonomic groups. People employed fire to modify vegetation and to concentrate game animals. They fished, hunted, and gathered in ways that affected species populations, animal behavior, and the dynamics of interspecific relationships. They domesticated plant and animal species, and altered natural hydrological systems through irrigation. In some cases—for example, the deforestation of Mediterranean watersheds and the development of hydraulic cultures in arid and semiarid lands—large-scale resource exploitation caused fundamental changes in biological diversity and ecosystem function. In other cases, the limited rate and scale of exploitation resulted in relatively minor long-term ecological impact.

Humanized landscapes, in sum, experienced varying degrees of anthropogenic environmental change. However, in many if not most native cultures, social mechanisms evolved to reinforce an attitude of respect and deference toward nature, to evoke nature's bounty, and to sanction appropriate human use of that bounty. These mechanisms included hunting, planting, and herding rituals; stories and myths; fertility and birth control practices; recognition of sacred spaces; and the invocation of taboos and totems. The conservation movement might be characterized as modern society's conscious effort to develop and exercise analogous social practices and restraints to guide its relations with the nonhuman world.

B. Historic Precedents

Contemporary environmental dilemmas have prompted modern scholars to reexamine conservation traditions in texts and stories from animist, Native American, Buddhist, Hindu, Islamic, Judeo-Christian, and other faith traditions. In the Judeo-Christian tradition, the biblical injunction to "fill the earth and subdue it, and have dominion over every living thing that moves upon the earth" has often been cited as providing the justification and rationale for centuries of environmental exploitation in the Western experience. Only recently have biblical scholars returned to these textual sources to identify alternative traditions of stewardship and respect for creation.

Since ancient times, observers have recorded instances of environmental deterioration due to human action. Plato, for example, compared the deforested mountains of Attica to the "bones of a wasted body . . . the richer and softer parts of the soil having fallen away, and the mere skeleton being left" (Dubos, 1980). Evidence of early conservation practices can be found in the history of both Western and non-Western cultures. These include efforts to protect particular species and special lands, to maintain populations of wild plants and animals, and to sustain the productivity of agroecosystems. Leopold (1933), for example, cited Marco Polo's narrative account of hunting restrictions and provisions for game bird populations in 13th century China as "the first clear record of a well-rounded system" of wildlife conservation.

European traditions of forestry and game-keeping date back to the Middle Ages and beyond. The establishment of royal game preserves and forests on the land estates of feudal Europe led to the development of customs and formalized laws regulating hunting and use of the forests. In the eleventh century, William the Conqueror set up in the newly conquered England "many deer preserves and also enacted laws/That whoever killed a hart or hind/Should be blinded./He also placed a ban on harts, also on boars" (Harrison, 1992). The high-handedness of the Norman conquerors in establishing privileged use of the forests gave rise to a venerable tradition of local resentment over centralized resource management. As codified in forest law, however, these measures allowed England's forests to survive and regenerate. On the continent, Germany and France in particular developed silvicultural systems and techniques that prevented wholesale destruction of the forest estate.

Europeans carried these "proto-conservation" traditions to their expanding colonial empires. The change in jurisdiction over natural resources, from native peoples to colonial governments, had profound implications for social systems and ecosystems alike. Colonial domination, coupled with rising populations and industrialization, disrupted traditional patterns of land tenure and resource use. In many regions, native peoples became increasingly alienated from their landscapes, while Western conservation ideas and practices were slow to adapt to the new environments.

In the New World, European colonists encountered a landscape of abundance that admitted profligate resource use. Unchecked resource exploitation was the norm as European settlement of North America proceeded. Despite this record, seeds of the later conservation movement were apparent in early efforts to protect wild game populations, forests, soils, and special natural features. Early measures to conserve game include decrees to protect the cahow (*Pterodroma cahow*) and green turtle (*Chelonia mydas*) in Bermuda (1616, 1622); the first closed hunting season on white-tailed deer (*Odocoileus virginianus*) in Massachusetts (1694); closed seasons on several species of game birds in New York (1708); and the institution of game warden systems in Massachusetts (1739) and New York (1741) (Matthiessen, 1959). The first federal game law, enacted in 1776, mandated closed seasons on deer in all the American colonies except Georgia. Through the 1800s the new American states intermittently passed laws establishing closed seasons, prohibiting hunting of nongame birds, and placing bounties on predators.

Exploitation of North America's extensive native forests—for conversion of land to agriculture, for construction materials, for charcoal and domestic fuel supplies—drove economic development through much of the colonial and early American era. In the late seventeenth and early eighteenth century, the British crown instituted the Broad Arrow policy, which reserved for the crown the right to cut the strategically important white pine (*Pinus strobus*) mast trees of the New England colonies. Although serving to restrict exploitation, these laws were intended more to secure the property against unauthorized timber cutting than to protect or perpetuate the forests per se. Closer in intent to later forest conservation measures was William Penn's 1681 dictum that, in the process of land development, one acre of forest should be maintained for every five acres cleared.

Little heed was paid to such recommendations. By 1745 Benjamin Franklin was noting the scarcity and high cost of fuelwood in Philadelphia. President James Madison in 1818 drew attention to the "injurious and excessive destruction of timber and firewood" as the most important and regrettable force shaping the rural economy of the United States. Such destruction continued through the 19th century. With the depletion of the eastern forests and the continuing settlement of interior North America, the focus of forest exploitation shifted to the Great Lakes region. In 1831 Alexis de Toqueville, while visiting Michigan, noted that Americans were generally "insensible to the wonders of inanimate nature, and may be said not to perceive the mighty forests that surround them till they fall beneath the hatchet." The swift destruction of the Great Lakes forests in fact marked a turning point in forest conservation, and in the conservation movement generally (Williams, 1989).

A parallel pattern of resource degradation and early conservation response marked the process of agricultural development in North America. By 1776 the problem of soil erosion was evident to the new nation's leading statesmen. Thomas Jefferson experimented with contour plowing to retard soil erosion. Patrick Henry, speaking before the Virginia Assembly, stated that "Since the achievement of our independence, he is the greatest patriot who stops the most gullies." George Washington noted in a 1797 letter, "We ruin the lands that are already cleared and either cut down more wood if we have it, or emigrate into the western country" (Jackson, 1985). Over the next century, destructive agricultural techniques continued to affect widely varied ecosystems across North America (and other parts of the world as well).

The preservation impulse in conservation found special expression in the North American setting, as European settlers encountered scenic landscapes and natural features unlike any in their prior experience. Jefferson celebrated Virginia's Natural Bridge in his *Notes on the State of Virginia.* Niagara Falls, the Hudson River valley, and the prairies, mountains, and canyons of the American West became emblematic of the beauty to be found in New World landscapes, a quality seized upon by artists, promoters, and early conservationists alike. The artist George Catlin suggested after his first excursions in the West that portions of the continent's undeveloped lands were "worthy of our preservation and protection." As early as 1833, Catlin envisioned government reservation of "a magnificent park ... a Nation's park, containing man and beast, in all the wild[ness] and freshness of their nature's beauty" (Nash, 1982). Such protected areas would not come into being until later in the nineteenth century, although the United States did move as early as 1832 to protect special natural features when it reserved Hot Springs in Arkansas as a federal holding.

What these disparate efforts had in common was not the conservation of biological diversity in the modern sense, but the protection or regulated exploitation of economically and aesthetically important components of the landscape or ecosystem. As such, they provided the foundation on which a more coherent conservation movement began to take shape in the latter half of the nineteenth century.

II. MODERN ORIGINS OF CONSERVATION

A. Prelude to a Movement

The conservation movement emerged in response to accelerated changes in demographic and environmental conditions, and in human perceptions and understanding of the natural world, through the eighteenth and nineteenth centuries. Exploration of the world's diverse ecosystems—from the earliest voyages of discovery through the New World scientific expeditions of Alexander von Humboldt, John and William Bartram, Meriwether Lewis and George Rogers Clark, John James Audubon, and others—contributed to a golden age of natural history studies. Establishment and adoption of the Linnean system of binomial nomenclature in the 1700s allowed for an unprecedented flourishing of taxonomic research (Mayr, 1982). This, in turn, provided critical foundations for the development of evolutionary theory in the work of Alfred Russell Wallace and Charles Darwin. Although the biological sciences had not yet developed field methods for comprehending the full diversity of life, they had begun to reveal the fundamental processes through which life in fact diversified and maintained itself.

These gains in systematics, biogeography, and evolutionary theory occurred even as the Industrial Revolution accelerated the scale, pace, and character of human environmental impacts. Through the 1800s, the advent of more efficient technologies resulted in increasingly intensive exploitation of forests, game populations, fisheries, agricultural lands, and river systems. Traditional resource management practices and established land tenure systems were abandoned or changed to fit the emerging economies of scale. Industrial pollution and the spread of invasive species became widespread problems for the first time.

Coincident with these scientific, cultural, and environmental changes, the Enlightenment and Romantic movements were altering Western conceptions of order, value, and beauty in the natural world. The natural philosophers of the Enlightenment stressed the smooth workings and stability of a mechanistic natural order. The Romantic philosophers and poets emphasized the unity and wholeness to be found in a spontaneously creative organic nature. Although offering different conceptions of nature, both encouraged human comprehension of natural objects and processes, and so laid the foundation for greater appreciation of human impacts upon the natural world. The writings of

Thomas Malthus and other early economic philosophers provided the basic framework for considering the interwoven fate of human population, human economies, and natural resources.

In Europe the Romantic movement drew heavily upon the experiences of New World explorers and settlers, the encounter with native peoples there and elsewhere, and the exposure to wild landscapes. In turn, adaptation of the Romantic impulse in the North American setting provided important literary and philosophical foundations for conservation. The strong American identification with wild nature founds its first full expression in the essays of the transcendentalists Ralph Waldo Emerson and Henry David Thoreau, the poems of William Cullen Bryant, and the novels of James Fenimore Cooper (Nash, 1982).

George Perkins Marsh's *Man and Nature: or, Physical Geography as Modified by Human Action* (1864) is widely regarded as the first landmark in modern conservation literature. A native Vermonter, Marsh saw in the destruction of New England's forests the latest expression of an ancient human tendency to "[derange] the original proportions between different orders of organic life." Drawing also on his extensive personal observations of long-term landscape change in the Mediterranean, Marsh argued that human actions had caused widespread disruption of the "harmonies" of the natural world. Marsh's reasoning followed lines that would sound familiar to later generations of ecologists and biodiversity conservationists. "All nature," he wrote, "is linked together by invisible bonds, and every organic creature, however low, however feeble, however dependent, is necessary to the well-being of some other among the myriad forms of life with which the Creator has peopled the earth" (Marsh, 1864).

B. Gaining Definition

Publication of Marsh's book provided direction to the conservation movement as it gained greater definition through the remainder of the 1800s and the first decade of the 1900s. In North America, dispossession of the American Indians, enactment of liberal land distribution policies, and the flow of settlers and capital into "virgin" landscapes resulted in an unprecedented wave of exploitation of natural resources. The effects were visible in ecosystems across the continent. These decades saw the virtual depletion of the extensive pine forests of the upper Great Lakes; overhunting of many game bird and mammal species by market hunters; rapid conversion of the Midwest's extensive prairies to agriculture; degradation of aquatic systems through overfishing, pollution, and hydrological changes; and widespread overgrazing of the semiarid western rangelands. Several well-recognized cases of species depletion and extinction—of the white pine, of waterfowl, of the bison (*Bison bison*), the Carolina parakeet (*Conuropsis carolinensis*), and the passenger pigeon (*Ectopistes migratorius*)—came to symbolize this era of unmitigated resource extraction and decline.

This period, however, also saw the first concerted efforts to address the causes and consequences of these massive changes. At first these actions tended to focus on particular issues, problems, or landscapes. Sportsmen led the campaign to rein in market hunting and to institute stronger game laws at the state and local levels. Many of the nation's most prominent sportsmen banded together in 1887 to form the Boone and Crockett Club, which became one of the first and most effective nongovernmental organizations to become involved in shaping national conservation policy. Exploitation of plume-bearing egrets and other birds for the millinery trade mobilized sportsmen, scientists, and nature lovers to work together for reform, and prompted formation of state Audubon Societies beginning in 1896. These moves culminated in passage of the Lacey Act of 1900. The first important piece of modern wildlife conservation legislation, it barred interstate shipment of wild animal species taken in violation of state laws, and soon succeeded in ending the plume trade (Dunlap, 1988)

Efforts to reform American land policy proved more difficult. Through the 1800s, the nation's land allotment policies had encouraged rapid immigration, settlement, and land conversion, often accompanied by corruption and land speculation. In the midcontinent's mesic landscapes, the system of land surveying and distribution facilitated the conversion of millions of acres of native savanna and grassland to crop and livestock production. These policies, however, were less tenable in the progressively drier lands of the American West, where farming was possible only near permanent watercourses, and livestock ranching required large land holdings. The geologist and explorer John Wesley Powell attempted in the 1870s and 1880s to devise a more appropriate tenure system that recognized the inherent environmental limitations and social requirements of the landscape, but the speculative momentum proved intractable. Only decades later did the U.S. government institute more careful land management policies on the nation's remaining public domain. However, Powell's innovations, especially his commitment to a strong role for science in land management, laid important foundations for future private and public conservation planning (Stegner, 1954).

The rapid destruction of forests in the upper Great Lakes between 1865 and 1900 stimulated the most forceful and effective response among early conservation reformers. As this period began, forestry as a profession had not yet gained a foothold in the United States. Through the 1870s and 1880s, however, voices opposed to the destructive "cut and run" logging practices began to emerge. These reformers included botanists Charles S. Sargent, Asa Gray, and William H. Brewer; naturalists John Muir and Increase Lapham; the European-trained forester Bernard Fernow; Secretary of Interior Carl Schurz; and Franklin B. Hough, appointed as the nation's first government forestry official in 1876.

The forestry movement gained momentum quickly in the last decade of the 19th century. Forest activists working through professional scientific and forestry organizations campaigned for national legislation to reform public land laws and to prevent further forest destruction. These efforts culminated in 1891 with the adoption of legislation that included a provision allowing the U.S. president to set aside public land forests as forest reserves, the germ of the nation's current system of national forests. Within weeks of its passage, President Benjamin Harrison used the law to designate forest reserves in lands adjacent to Yellowstone National Park. In the next 10 years, 47 million acres of forest land were placed in reserves.

During this period, the loss of forests contributed also to the parallel movement to protect and preserve unusually scenic landscapes, special natural features, and wild spaces. Like the forestry movement, this channel of activism gained momentum in the decades following the American Civil War. In 1872 the U.S. Congress established Yellowstone National Park, the world's first. In 1885 New York created a state forest preserve in the Adirondack Mountains in order to protect its wild character as well as its watershed values. That same year Canada designated Banff National Park, its first. The impetus toward preservation drew heavily upon the enthusiastic nature writing of John Muir, who led the political effort that in 1890 resulted in designation of an enlarged Yosemite National Park in California and who in 1892 founded the Sierra Club (Fox, 1981).

At the end of the 19th century, the varied strands of conservation concern—over depleted wildlife, unprecedented deforestation, watershed degradation, inappropriate land development, loss of scenic values, and aesthetic quality—were connected only loosely. In the American experience, political corruption and inordinately concentrated wealth had typically accompanied these environmental changes. As the new century began, these overlapping social, economic, political, and environmental concerns brought forth a more consolidated movement that gave new meaning and power to the term *conservation*. However, the growing prominence of conservation also revealed inherent tensions in the nascent movement.

C. Progressive Era Conservation and the Utilitarian/Preservationist Split

When Theodore Roosevelt assumed the American presidency in 1901, the stage was set for a revolution in conservation policy. Roosevelt brought to the presidency a formidable command of the natural sciences, experience as a rancher and outdoorsman in the American West, a long-standing devotion to public policy reform, and boundless personal energy. In his first presidential address to the U.S. Congress, Roosevelt spoke at length of the importance of the nation's forests, stating that "We have come to see clearly that whatever destroys the forest ... threatens our well-being" (Pinchot, 1947). Conservation became a cornerstone of the Progressive political movement and, for the first time in history, a high national priority.

Roosevelt's partner in political innovation was forester Gifford Pinchot. Pinchot, the first American to receive formal training in forestry, was a friend of Roosevelt's who had been active in the complex politics of American forestry through the 1890s. At the time Roosevelt became president, Pinchot was in charge of the Division of Forestry in the U.S. Department of Agriculture. The two quickly forged a close alliance to strengthen the government's role in forestry. Their partnership during Roosevelt's presidency resulted in an increase in the total acreage of the forest reserves (renamed *national forests* in 1905) from 60 million to 151 million acres; transfer of these lands to the Department of Agriculture under Pinchot's jurisdiction; and creation, in 1905, of the U.S. Forest Service to administer the new national forests.

Pinchot's Forest Service embodied the Progressive spirit and approach to bureaucratic responsibilities. In contrast to Muir and others in the preservationist wing of the conservation movement, Pinchot conceived of the forests in utilitarian terms. The forests were not regarded as "reserves" to be "locked up," but as lands to be worked "for the greatest good of the greatest number for the longest time" (as the utilitarian credo put it). In practice, this meant that the forests were to be managed by a trained, professional workforce; that scientific principles were to guide the efficient and sustained exploitation and processing of forest resources; and that the wealth derived from the forests was to be equitably distributed for the common good (Hays, 1959). Applied not only to forests, but to natural resources in general, the "resource conservation ethic" provided the dominant paradigm of the early movement (Callicott, 1991). In Pinchot's words, "The first great fact of conservation is that it stands for development" (Pinchot, 1910).

At the core of utilitarian conservation was the concept of sustained yield. As Aldo Leopold later (1933) observed, under the Progressive conservation banner "wild life, forests, ranges, and waterpower were conceived ... to be *renewable organic* resources, which might last forever if they were *harvested scientifically, and not faster than they reproduced.* 'Conservation' had until then been a lowly word, sleeping obscurely in the dictionary. The public never heard of it. It carried no connotation of woods or waters. Overnight it became the label of a national issue." As the leading edge of this movement, the Forest Service became a model not only for other resource management agencies but for Progressive-era government agencies in general.

Theodore Roosevelt's conservation activism extended beyond the public forests. During his presidency he established the nation's first wildlife refuge, at Florida's Pelican Island (1903); designated 50 additional wildlife refuges; used new presidential powers under the Antiquities Act of 1906 to set aside 18 national monuments (including the Grand Canyon); convened in 1908 a national Governor's Conference on the Conservation of Natural Resources; and created, also in 1908, a National Conservation Commission to provide continuing advice to the president and to work with the states on conservation policy. Early in 1909, Roosevelt and Pinchot sought to broaden the scope of conservation through a North American Conservation Conference and laid still more ambitious plans for a World Conservation Conference. Though the latter was unrealized, Pinchot and Roosevelt had succeeded in securing the place of conservation within both the political and geographic landscape.

Absent from many of the Progressive-era political initiatives were the voices of the increasingly influential preservationists and nature protectors. In contrast to the utilitarian views that Pinchot exemplified, adherents of the "romantic-transcendental preservation ethic" emphasized wild nature's aesthetic and spiritual values and the need to safeguard those values for future generations through strict prohibitions on development and manipulation (Callicott, 1991). The tensions between the utilitarian and preservationist approaches intensified as conservation assumed center stage politically. These tensions surfaced in the changing relationships among the principal players. Pinchot and John Muir had been friends, but by the late 1890s their differing approaches to management of the forest reserves caused a rift between them that would never heal. Few from the preservation camp participated in the landmark 1908 governors'conference on conservation.

Muir maintained a respectful relationship with Roosevelt. They camped together in Yosemite in 1903 and worked together in gaining protection for the Grand Canyon as a national monument in 1906. But the schism between the conservation factions continued to widen. It came to a head in the bitter political struggle over plans to dam the Tuolumne River in Yosemite National Park's Hetch Hetchy Valley. Waged over a period of 6 years, the battle culminated in 1913 with the adoption of national legislation providing federal support for the dam. Although Muir and his colleagues lost the battle, they had aroused a national constituency in favor of protection. The growing popular and political acceptance of preservation led directly to creation of the U.S. National Park Service in 1916 (Fox, 1981).

For all of the profound developments in conservation during the Progressive Era, scant attention had as yet been given to the state of the biotic diversity and ecological processes characteristic of either wild or more humanized landscapes. That attention would come only slowly, as conservation science, philosophy, policy, and practice co-evolved through the twentieth century.

III. THE EVOLUTION OF TWENTIETH-CENTURY CONSERVATION

The Progressive era advances fixed conservation permanently in the public mind and in public policy. The innovations of that period, however, served only to suggest the basic questions that would challenge conservationists throughout the twentieth century: What are we conserving? To what end? For whom? How? With what scientific information and understanding? With what social adjustments, economic mechanisms, and legal tools? Moreover, the conservation movement came out of the Progressive Era with the tension between the preservationist and utilitarian approaches still unresolved. The continuing efforts to answer those questions and to resolve that tension framed the century's technical and political advances in conservation.

A. Institutionalizing Conservation

In the United States, the first four decades of the twentieth century saw the consolidation of the Progressive era's conservation gains in the government, the private sector, the professions, and the academy. The spirit of reform that drove the conservation movement was now redirected toward the development of sound conservation policy and administration. New issues, however, arose to keep the movement invigorated and to expand its philosophical dimensions.

As conservation became institutionalized, it tended to follow the tenets of Pinchot's resource conservation ethic. By the late 1930s, the basic principles of utilitarian resource conservation had been applied not only to forests, but to other "useful" components of the biota and the landscape: rangelands, game animals, sport and commercial fisheries, scenic areas, agricultural soils, river systems. New policies, laws, bureaucracies, academic disciplines, research and training programs, and professional societies arose to promote sustained yields of and from these various "resources." Foresters, range managers, wildlife managers, recreation planners, and soil conservationists achieved their own professional identities in these years. Amateur conservationists, meanwhile, became increasingly involved in decision making through the formation of new citizen organizations.

Forestry. As an established field, forestry provided the model and foundation for the other emerging resource management professions in the United States. With those fields still lacking formal training and employment opportunities, forestry tended to attract professionals with broad conservation interests. However, institutionalization tended over time to narrow the scope of the field to silviculture and timber harvesting. This trend was magnified during these decades as the supply of timber taken from private lands diminished and the pressures on public forests increased.

Range management. In the American west, rapid settlement and exploitation had degraded vast stretches of arid and semiarid grassland in the late 1800s and early 1900s. Research and academic training programs followed in the wake of the deteriorating range conditions, bringing scientific management to rangelands much as forestry had brought such management to forests. Changes in land policy resulted by the late 1930s in stricter regulation of the public rangelands, cessation of privatization of the public domain, establishment of a new federal bureau (the Bureau of Land Management) to oversee these lands, and the emergence of range management as a discrete field.

Wildlife management. Although concern about the fate of wild animal populations had often galvanized the young conservation movement, wildlife management did not carve out its own professional niche until the 1930s. Continuing declines in game populations provided the initial impetus. To arrest these declines, game protectors at first focused on enactment of stricter hunting regulations, sporadic establishment of refuges, development of artificial propagation programs, control of predators, and the introduction of exotic game species. Only in the early 1930s did a revolutionary new approach, focused on the protection and restoration of habitats so that populations might sustain themselves, begin to take hold. (Fisheries management followed a roughly parallel pattern with regard to game fish species and habitats.) As principles from ecology filtered into the field, the term *wildlife* came into common usage, and the idea of management expanded to encompass the perpetuation of "nongame," rare, and threatened species. Although the field tended to focus primarily on game species, it provided important foundations for efforts later in the century to conserve biological diversity generally.

Recreation. As the American population became increasingly urban and increasingly mobile, outdoor recreation became a more important component of conservation. Camping, fishing, hunting, hiking, and other outdoor activities continued to grow in popularity, even after the prosperity of the 1920s gave way to the Great Depression of the 1930s. Park agencies, at all levels of government, responded by establishing and developing areas of significant aesthetic and recreational value. Other land management agencies, such as the U.S. Forest Service, began to devote greater attention to the recreational values of the lands under their jurisdiction. Although there were exceptions, recreational developers tended to override or ignore problems associated with the protection or management of the biota, and they rarely incorporated scientific information in planning for increased tourism.

Soil conservation. The expansion and mechanization of agriculture in the early decades of the twentieth century brought widespread disruption to the agricultural landscape and economy. The disaster of the Dust Bowl in the southern Great Plains during the mid-1930s was the most highly visible consequence of forces that had degraded soils across North

America. Intensified exploitation of farm resources through land clearing, "clean farming," poor grazing management, and wetland drainage led to accelerated rates of soil erosion, degradation of pasture and forage vegetation, loss of wildlife habitat, and aggravated cycles of siltation and flooding. In response, the U.S. Congress in 1935 created the Soil Conservation Service (SCS) to work with farmers and other private landowners to adopt conservation practices. Under the leadership of Hugh Hammond Bennett, the SCS importantly emphasized watershed-scale approaches that integrated soil protection measures with other resource conservation goals.

Although increasingly defined by this enhanced structure of agencies, disciplines, and professions, the conservation movement remained a battleground for competing approaches and intense political activity. Even as conservation became professionalized, growing numbers of citizens became active as members of nongovernmental conservation organizations. New national groups, including the Izaak Walton League, the National Wildlife Federation, and Ducks Unlimited, joined such older organizations as the Sierra Club and the Audubon Association as significant players in shaping conservation policy (Fox, 1981).

Meanwhile, the preservationist approach found renewed vigor as the campaign to protect wildlife, natural areas, and wildlands reemerged in the 1920s and 1930s. In these years, field zoologists such as Joseph Grinnell and Olaus Murie became leaders in the effort to reverse government antipredator policies. Early ecologists, under the leadership of Victor Shelford, created in 1917 a Committee on the Preservation of Natural Conditions within the Ecological Society of America. This committee provided the nucleus for what became, decades later (1950), The Nature Conservancy. In 1924, the U.S. Forest Service designated the nation's first "wilderness area" on a large roadless expanse of the Gila National Forest in New Mexico. In 1935, proponents of wilderness protection in the United States banded together as The Wilderness Society, the first conservation organization dedicated solely to the cause of wildland protection. Although the wilderness protection movement still emphasized primarily the aesthetic, spiritual, and recreational value of wildlands, these moves signaled a shift toward greater recognition of the scientific and ecological value of undeveloped lands.

In the international arena, the years prior to World War II saw increased recognition of the global scope of conservation challenges and halting moves to institutionalize a response to those challenges. The western powers—particularly the United States, Britain, France, and Germany—continued to export forestry and national park programs to their colonies and protectorates. Early international conventions and treaties addressed the problems of protecting fur seals, sea otters, whales and other marine species, and migratory birds (Matthiessen, 1959). In 1913 an International Conference on the Protection of Nature in Switzerland attracted representatives from 16 nations. World War I prevented this effort from gathering momentum, but in the 1920s and 1930s intermittent steps were taken toward greater international cooperation: a second international conservation conference in Paris in 1923; establishment in 1928 of an International Office for the Protection of Nature in Brussels; an unprecedented international bird conservation conference in Paris in 1933; a Conference for the Protection of African Fauna and Flora in London, also in 1933; the first North American Wildlife Conference in 1936; establishment of the Pan-American Union, dedicated largely to conservation issues in the Western hemisphere, in 1940 (Nash, 1982). These set the stage for the expanded international programs that emerged in the aftermath of World War II.

B. Conservation Amid Crisis: The 1930s and 1940s

The social and environmental convulsions of the 1930s and 1940s fundamentally altered perspectives and priorities within the conservation movement. The Depression forced conservationists and nonconservationists alike to consider the connections between human economic systems and the sources of wealth in nature. In practical terms, the New Deal programs in the United States vastly increased governmental support for conservation programs, though often with little consideration of ultimate conservation goals. The experience of the Dust Bowl focused the attentions of an entire generation of conservationists and many natural scientists on the consequences of indiscriminate exploitation of sensitive lands, highlighting especially the systemic nature of the soil erosion problem. With the outbreak of World War II, conservation issues fell into the background of concerns, proving the difficulty of maintaining mindfulness of human-nature relationships when social crises erupt. As in no time since the Progressive Era, these years of crisis demonstrated the interwoven nature of social, economic, and environmental problems.

As a movement, conservation remained prominent in North America and elsewhere, with growing popular and political acceptance. But for all of the recent advances, the conservation movement still suffered from its own internal crisis. Beneath the very active surface of conservation, there remained the underlying philosophical rift between the utilitarian and preservationist approaches. Neither approach adequately addressed such extensive problems as soil erosion and exhaustion, disruption of hydrological cycles, endangerment and loss of wildlife species, degradation of biotic communities (both terrestrial and aquatic), and the increased incidence of invasion by exotic pests and diseases. For a small but growing group of conservationists, the Dust Bowl (among other factors) stimulated a new effort to address the roots of conservation problems, one that would draw on elements of both the utilitarian and preservationist schools. Embracing the preservationist critique of human hubris in manipulating the natural world, the aesthetic appreciation of wild nature, the traditional commitment to wise use and stewardship of resources, and the value of science in the service of conservation, this cohort of conservationists sought to define a more robust synthesis to guide conservation action.

The incipient synthesis reflected a basic shift in conservation's scientific foundations. Theory and application—basic science and conservation practice—were meeting in ways they had not before. Developments in ecology and evolutionary biology over the previous decades had began to inform conservation issues, even as contemporary problems forced conservationists to reexamine their scientific assumptions. Ecology was revolutionizing scientific understanding of the functioning of biological communities, landscapes, and systems. Evolutionary biology provided new perspectives on, for example, the adaptations, roles, and interactions of all forest species (in contrast to the basic descriptive botany, dendrology, timber physics, and forest mensuration upon which silviculture and forestry had previously rested).

For at least some conservationists, these scientific advances suggested a new need: to marry ecology and the various fields of resource management in the effort to sustain not just yields of particular commodities but the healthy functioning of the systems generally. The preeminent voice of this emerging approach was Aldo Leopold, who in the course of his career had applied ecological principles first to the conservation of forests, then to soils, game, and wildlife species, then ultimately to the land "as a whole." Writing in 1939, Leopold noted that ecology provided "a new fusion point for all the natural sciences" and that its emergence had

> placed the economic biologist in a peculiar dilemma: with one hand he points out the accumulated findings of his search for utility, or lack of utility, in this or that species; with the other he lifts the veil from a biota so complex, so conditioned by interwoven cooperations and competitions, that no man can say where utility begins or ends .... The only sure conclusion is that the biota as a whole is useful, and [the] biota includes not only plants and animals, but soils and waters as well (Leopold, 1939, p. 727.)

Leopold's expanded conservation philosophy, as finally expressed in his landmark essay "The Land Ethic" in *A Sand County Almanac* (1949), placed greater emphasis on the ecological integrity, diversity, and beauty of what he called "the biotic community" and rejected the view of nature as merely a collection of disaggregated natural resources. It shifted the role of human beings "from conqueror of the land community to plain member and citizen of it" (Leopold, 1949). This proposed shift, away from an anthropocentric toward a more biocentric approach, implied the need for fundamental changes within the various fields of resource management. Resource managers could no longer regard timber trees, game animals, water, soils, scenic vistas, or any other economically or aesthetically significant resource as separate entities. Rather, resources had to be seen as components within diverse systems, connected to and interacting in complex ways with other parts of the system (including human beings). It followed that the conservation professions had to develop more integrated approaches to resource management.

As an attempt to address conservation's underlying schism, Leopold's "evolutionary-ecological land ethic" would confront continuing challenges in the dramatically altered postwar environment (Callicott, 1991). However, it would also provide those in the conservation movement with new conceptual tools with which to meet those challenges.

C. From Conservation to Environmentalism

The conservation movement emerged from World War II into a changed world. The war altered conservation's social and ecological context. In the world at large, the bonds of empire dissolved in the postwar decades as colonial nations gained their independence. The former colonial powers remained involved in the fate of their erstwhile colonies as donors and providers of development assistance. This in turn altered the shape of the global conservation movement, as aide providers, international agencies, and nongovernmental agencies initiated the long struggle to somehow reconcile development and conservation priorities.

These changes in international relations occurred even as human population growth, land degradation, air and water pollution, and overexploitation of marine resources first emerged as global-scale conservation issues. The war had demonstrated the interrelated and global nature of modern conservation problems. Fairfield Osborn's *Our Plundered Planet* (1948) and William Vogt's *The Road to Survival* (1948) were among the first books to address in an integrated manner issues of global conservation and development. Reflecting this broader scale of concern, a wide spectrum of national and international governments and organizations met in 1948 and formed the International Union for the Protection of Nature (now known as the World Conservation Union) with the ambitious goal of preserving "the entire world biotic environment." Shortly thereafter, the United Nations became involved in international conservation programs. These were early indicators of what became an increasingly active international response to conservation problems.

In the United States, the war years constituted a demographic divide, with important implications for the conservation movement. As veterans and their families resumed their lives at home, the postwar population boom fueled an economic boom that placed increasing demands on natural resources. To meet those demands, war-spawned technologies and industrial processes were quickly adapted for use in the peacetime economy. Over the next several decades, rings of suburbs expanded away from cities into rural lands and, along with dams, highways, and other infrastructural developments, continually reconfigured the landscape.

The postwar years brought a new generation of professionals into agriculture, forestry, wildlife management, and the other conservation fields. Conservation was not immune to the larger social and economic forces in the postwar environment. As the pressure to meet rising demands increased, the resource management professions became more specialized, more focused on commodity outputs, and more inclined to adopt technical solutions to conservation problems. Importantly, the war also changed the ways in which scientific research was supported and conducted, reinforcing the tendency toward specialization and utilitarian applications. The traditional foundations of conservation in systematics, organismal biology, population biology, ethology, and other field-oriented sciences suffered from relative neglect. Collectively these changes overwhelmed the integrated approach that had begun to emerge prior to the war, requiring conservationists to confront complex environmental problems within an increasingly rigid framework of disciplines and institutions.

However, the postwar years also gave rise to forces that countered these trends. New tools in the earth sciences provided greater scientific understanding of the interrelations within and among terrestrial, aquatic, and atmospheric systems. In the 1950s and 1960s, revolutions in fields ranging from genetics and evolutionary biology to atmospheric chemistry and geology began rapidly to reshape our understanding of the global biosphere and the human place within it. New communications technologies allowed such information to be more efficiently shared. Meanwhile, the indiscriminate adoption of other technologies—especially in the production of agricultural pesticides, industrial chemicals, and nuclear power—created new concerns over their impacts on human health and ecological systems. With the publication in 1962 of Rachel Carson's landmark book *Silent Spring*, examining the biological impacts of DDT and other pesticides, the modern environmental movement began to assume an identity of its own, distinct from but still connected to the older conservation tradition.

Efforts to protect wildlands in the United States also intensified in the years following World War II. As pressures from urban and suburban expansion, resource extraction, water development, and motorized tourism increased, wilderness protection became a unifying goal among diverse conservation organizations. In particular, the defeat in 1955 of the proposal to build the Echo Park Dam in Dinosaur National Monument in Colorado indicated that the conservation movement had gained not only in cohesion, but in popular and political power. That unity led to the passage in 1964 of the Wilderness Act, providing for a strengthened national system of wilderness areas on the public lands of the United States.

The Wilderness Act was only one of many laws that emerged from this period of environmental awakening (much as earlier laws had institutionalized the conservation movement). Over the next decade, the U.S. Congress would enact a series of important environmental statutes, including the National Environmental Policy Act (1970), the Clean Air Act (1970), the Clean Water Act (1972), and the Endangered Species Act (1973). These far-reaching changes in environmental policy reflected a rising wave of popular support and organized political activism, symbolized by the observance of the first Earth Day on April 22, 1970.

D. The Reintegration of Conservation

With the emergence of the environmental movement, the context of conservation once again experienced dramatic changes. The years following Earth Day saw increasing acceptance, within the conservation professions and within society at large, of environmental values. Those values, however, were not easily translated into effective conservation action. The resource management professions, in seeking to address broader environmental concerns, found the tendency toward overspecialization difficult to overcome. Conservation programs tended still to focus on single species, or particular economic resources, or separate jurisdictions within a given landscape. As the long-term ecological impacts of such fragmented approaches to resource management became clear, the traditional conservation fields came under increasing public scrutiny. While many environmentalists pressed for change through legal means, the evolution of management philosophy, policy, and practice would require years of incremental change and adjustment.

At the international level, differences in perspective between the wealthier, developed countries of the North and the poorer, developing countries of the South likewise proved difficult to overcome. International conservation continued to make important gains through a series of conventions and treaties, including the Convention on Wetlands of International Importance (the "Ramsar Convention") (1971), the Convention on International Trade in Endangered Species (CITES) (1975), the International Convention for the Conservation of Migratory Species of Wild Animals (the "Bonn Convention") (1978), the United Nations Convention on the Law of the Sea (1982), and the United Nations Convention on Biological Diversity (1992). Again, however, international development policies would come to incorporate stronger conservation and environmental provisions only gradually, as the need to connect economic development and long-term environmental security became increasingly evident.

Even as these broad patterns of change in the human dimensions of conservation unfolded in the 1970s and the 1980s, the scientific foundations of conservation were shifting. Taxonomy and systematics provided more robust estimates of the extent of species diversity and of its actual and potential loss. Advances in island biogeography revealed the principles governing the spatial distribution, persistence, and extinction of species, with important implications for the establishment and management of protected areas. Genetics became an increasingly important component of conservation science as attention focused on the reproductive success of rare and endangered species and the viability of their populations, both in captivity and in the wild. Ecology moved away from its "classical paradigm," which emphasized single, stable, deterministic equilibria, and toward a view of ecosystems that emphasized flux, uncertainty, and contingency. Disturbances such as fire, flooding, and drought were incorporated into our understanding of long-term ecosystem dynamics and were shown to be important factors affecting species diversity. Increasingly, conservation strategies required the integration of knowledge from the many branches of biological science, involving various levels of biological organization.

The need to rethink conservation across disciplinary lines was driven not only by changes in the foundational sciences, but by changes in the environment itself. By the late 1970s, scientists and conservationists were alarmed by the escalating loss of genetic, species, and ecosystem diversity at the global scale. Of special concern was the accelerated destruction of the species-rich forests of the humid tropics. The advances in island biogeography revealed that ecosystems of all types were being fragmented by human activity, while even the most effective protected areas were at risk due to their inadequate size and their isolation. Wildland managers increasingly understood that preservation alone was an inadequate management strategy, and that the loss of diversity and the disruption of ecological functions were intimately associated. Agricultural scientists, foresters, and other resource managers, too, were increasingly concerned about environmental degradation, the breakdown of ecosystem processes, the loss of diversity in cultivated landscapes, and the human costs associated with these changes.

These concerns prompted the emergence in the mid-1980s of a new synthetic discipline, conservation biology, specifically devoted to the integration of scientific disciplines and other fields in the effort to understand, maintain, and restore biological diversity. To this end, conservation biology has sought to address conservation problems within an evolutionary and ecology context, and to stimulate the traditional conservation professions to reassess their management methods and goals accordingly. The term *biodiversity* itself was coined in 1987 and has since been widely adopted by conservationists. The concepts of *sustainability* and *sustainable development* came into general usage during the same period, reflecting the complex challenge of integrating long-term social, economic, and environmental factors in assessing human demands and impacts on ecosystems. The integrating impulse was also evident in the recent emergence of restoration ecology, landscape ecology, sustainable agriculture, environmental history, environmental ethics, ecological economics, and other interdisciplinary fields.

These recent changes in conservation science, philosophy, and policy have prompted many in the traditional resource management fields to adopt more integrated approaches. In practical terms, this means that as the conservation professions approached the new century, they sought to move away from narrow economic criteria for success and toward broader ecological standards; away from the mathematical *desideratum* of sustained yield of commodities and toward the more complex challenge of sustaining healthy ecosystems; away from the management of discrete components of the system and toward integrated ecosystem management; and away from an exclusive focus on the human goods and services provided by ecosystems and toward a focus on the biological diversity and ecological processes that sustain those goods and services. These changes have also begun to redefine the value of wilderness and to reintegrate wildlands within the broader landscape of conservation. Whereas wilderness as traditionally defined existed apart from the humanized landscape (even while being preserved to meet human recreational and aesthetic needs), wildlands at all spatial scales have become increasingly valued as repositories of biological diversity, as core protected areas and corridors within greater ecosystems, and as "controls" against which to compare the human impact on more intensively utilized lands.

These examples of increasing integration in conservation hearken back in many ways to the ecology-based synthesis that Aldo Leopold and others sought to articulate prior to the onset of the Cold War. Since then, the science of ecology has evolved, analytical tools and technologies have grown vastly more sophisticated, environmental threats have intensified, and the social and economic context of conservation has changed dramatically. Nevertheless, there is continuity between contemporary efforts to expand conservation's cultural and natural connections and Leopold's efforts to "enlarge the boundaries of the community to include soils, waters, plants, and animals" (Leopold, 1949).

IV. EMERGING THEMES IN CONSERVATION

As in the past, conservation remains responsive to new demands, new information, and new realities. The sweeping changes that have reshaped the conservation movement since the mid-1980s will continue to alter both its human dimensions and its practical field methods. Although we are still too much embedded in the present to recognize all these recent developments, we can identify important contrasts between emerging themes and conservation's past practices.

In terms of its scientific foundations, conservation now seeks greater collaboration among the sciences, even while reaching beyond the sciences to build connections with philosophy, economics, history, literature, and the arts. Ecology continues to be the "fusion point" of the conservation-related sciences. In its conservation applications, ecology is bringing attention to a broader spectrum of species and to the processes that maintain diversity. To improve its chances of success, conservation will need to persevere in its efforts to marry ecology to the resource management professions, and to do so at various spatial and temporal scales. The integrating fields of conservation biology, landscape ecology, restoration ecology, sustainable agriculture, and ecological economics will need to resist the tendency toward specialization and reductionism to which even interdisciplinary fields are prone.

Emerging threats are making the conservationist's task even more precarious. These include the rapid spread of invasive exotic species (especially as the human economy continues to globalize), the uncertain impacts of global climate change, the increased presence of genetically modified organisms in the landscape, and the seemingly relentless fragmentation of habitats and ecosystems. At the same time, the sciences have recently been called upon to illuminate the connections among biodiversity conservation, human health, and the well-being of human communities. Accordingly, the active weaving of conservation science into community-based ecological restoration programs promises to redefine the role and methodology of science in conservation.

The conservation movement continues to be shaped by overarching social and demographic trends. The growth in human population and in resource consumption rates remains a profound factor determining conservation's future. To better inform that future, conservationists are seeking novel ways to integrate sustainable economies with effective protection, management, and restoration programs. In so doing, conservation planning efforts are expanding their reach across jurisdictional boundaries and across the landscape, recognizing the connections among wildlands, semiwild lands, "working" landscapes, and suburban and urban environments. Such cross-boundary forays have few precedents in conservation history and represent an important departure from the past.

In the past, conservation has also neglected human and cultural relationships to its own detriment. Increasingly conservationists recognize the need to integrate natural science, cultural traditions, and social relationships in the effort to protect and restore particular places. In seeking a better fit between the built environment and surrounding landscapes, conservationists have also begun to work more closely with architects, landscape architects, planners, and engineers to incorporate the emerging principles of conservation biology.

Meanwhile, the conservation movement itself continues to increase in diversity as individuals from varied backgrounds come into conservation as both professionals and amateurs. Local nongovernmental organizations are now regarded as key conservation players, and their role in conservation continues to grow. Although action at the international level remains necessary in addressing global threats to biodiversity and in supporting local initiatives, the assumption of greater local responsibility for ecosystem health is increasingly urgent. Community-based conservation projects, watershed restoration programs, land trusts, and other mechanisms are emerging to meet that need. As these experimental programs continue, they seek to create not just an enhanced conservation movement, but an enduring culture of conservation. In so doing, they extend the long history of increasing citizen involvement in conservation.

See Also the Following Articles

ECOSYSTEM, CONCEPT OF • ENVIRONMENTAL ETHICS • HISTORICAL AWARENESS OF BIODIVERSITY • INDIGENOUS PEOPLES, BIODIVERSITY AND • LITERARY PERSPECTIVES ON BIODIVERSITY • RELIGIOUS TRADITIONS AND BIODIVERSITY • STEWARDSHIP

Bibliography

Callicott, J. B. (1991). Whither conservation ethics? *Conserv. Biol.* 4, 15–20.

Dubos, R. (1980). *The wooing of earth.* New York: Charles Scribner's Sons.

Dunlap, T. (1988). *Saving America's wildlife.* Princeton, NJ: Princeton University Press.

Fox, S. (1981). *John Muir and his legacy: The American conservation movement.* Boston: Little, Brown.

Harrison, R. P. (1992). *Forests: The shadow of civilization.* Chicago: University of Chicago Press.

Hays, S. P. (1959). *Conservation and the gospel of efficiency: The progressive conservation movement, 1890–1920.* Cambridge, MA: Harvard University Press.

Jackson, W. (1985). *New roots for agriculture.* Lincoln and London: University of Nebraska Press.

Leopold, A. (1933). *Game management.* New York: Charles Scribner's Sons.

Leopold, A. (1939). A biotic view of land. *Journ. Forestry* 37, 727–730.

Leopold, A. (1949). *A Sand County almanac and sketches here and there.* New York: Oxford University Press.

MacPhee, R. D. E. (Ed.). (1999). *Extinctions in near time: Contexts, causes, and consequences.* New York: Plenum Press.

Marsh, G. P. (1864). *Man and nature: or physical geography as modified by human action.* Cambridge: Cambridge University Press, 1965.

Matthiessen, P. (1959). *Wildlife in America.* New York: The Viking Press.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution and inheritance.* Cambridge, MA: Harvard University Press.

Nash, R. (1982). *Wilderness and the American mind* (3rd ed.). New Haven and London: Yale University Press.

Pinchot, G. (1947). *Breaking new ground.* New York: Harcourt, Brace, and Co.

Pinchot, G. (1910). *The fight for conservation.* Seattle and London: University of Washington Press, 1967.

Stegner, W. (1954). *Beyond the hundredth meridian: John Wesley Powell and the second opening of the West.* Boston: Houghton Mifflin.

Williams, M. (1989). *Americans and their forests: A historical geography.* Cambridge: Cambridge University Press.

CROP IMPROVEMENT AND BIODIVERSITY

Giorgini Augusto Venturieri
Federal University of Santa Catarina

GLOSSARY

barreais Soils naturally enriched with organic matter and clay carried by the rain and/or a river.

caboclos Cultural and/or racial crossing between European and indigenous populations.

macaxeira Group of varieties of manioc (*Manihot esculenta* Crantz) with low levels of cyanidric acid. These varieties may be consumed without prior treatment to lower their toxicity, as must be done with other varieties of wild manioc. Synonyms: sweet manioc and aipim.

maromba Elevated flooring made of wood, used in flooded areas for placing cattle or plants during river flood periods.

reproductive isolation When a derived population is not able, for some reason, to breed and obtain fertile progeny with the original population.

wild manioc Maniocs with a high level of cyanidric acid.

IN THIS ARTICLE, concepts are discussed relating to natural biodiversity used to form new crops and for the improvement of already domesticated species. The majority of examples come from underexploited plants of the Amazonian indigenous peoples, little known in the international scientific literature, but which are splendid examples of the capacity for breeding and selection of traditional communities. The article also gives curious reports of Amazonian Indian practices of management and interchange of germplasm from useful plants.

I. WHAT IS DOMESTICATION?

The evolution of a plant occurs when there is selection; this can be the result of human activity, which generally values attributes of interest to people in detriment to those necessary for the plant's own survival in the wild. When this process reaches high levels of genetic modification, to the point that human aid is necessary for survival, the plant is considered to have achieved the stage of being domesticated (Harlan, 1975). Useful plants occur at various stages of domestication that require different degrees of management and cultivation. Generally, the more domesticated the plant, the more demanding it is as to degree of management and cultivation, but the fact of a plant being managed or cultivated does not imply that it is domesticated. However, any plant that has been managed or cultivated for a long time is inevitably undergoing modification toward domestication because it is placed in environmental conditions improved by humans, who reduce or eliminate the usual processes of natural selection such as competition and predation. Domestication is related to the plant's genetic response, management, and cultivation and human activity in dealing with the plant (Harlan, 1975).

This article will deal principally with practices still used by Indians and caboclos of Amazonia who may directly (conscious selection) and indirectly (unconscious selection) be inducing modifications in natural plants in the direction of their domestication.

II. HOW DOES DOMESTICATION HAPPEN?

The forces of selection, resulting from management practices or from cultivation per se and as induced by farmers, may be consciously or unconsciously provoked (Harlan, 1975). In conscious selection, the farmer sees some interesting attribute in a plant and seeks to propagate it in some way, be it by sexual means (seeds) or asexually (generally by trying to induce a portion of stem to take root), thus forming a new population. If the characteristic selected has good inheritability—that is, it can be transferred from one generation to another—the result is an improved population. In unconscious selection, as the name indicates, this selection does not occur but the populations are modified due to some human intervention, generally in the environment or harvest system, which ends up benefiting some segment of a population to the detriment of the others, generating a selected population. The classic example of unconscious selection is that of rice, which upon being harvested has seeds in its panicle that do not easily fall from the plant. When these seeds are planted, they carry alleles that favor this characteristic, which is selected increasingly more favorably with each harvest. Seeds of domesticated rice today do not fall as easily from the panicle as do those of more primitive varieties (Harlan *et al.*, 1973).

The well-known "genetic drift" of the classic works on evolution is also very much associated with the process of domestication. Genetic drift occurs when a very small portion of a certain population is taken to another area that is isolated and there regenerates, but does so only with the descendants of the few individuals introduced therein (Fig. 1). Many times this process does not bring gains in productivity but does induce an increase in more or less stable varieties that end up forming the "local varieties." This process is very common among farmers who have the habit of collecting only a few seeds from a given plant for planting on their farms or villages, generally far from the original population and isolated by some environmental barrier. This may be a large river or, in the case of plants typical of disturbed environments such as the plants domesticated in farm plots, a climax forest in which such species do not occur.

Currently domesticated plants were some time ago wild plants which were in stages used, then managed, and finally cultivated (Clement, 1992a,b). The process of domestication, for the majority of species, took thousands of years. For corn, whose domestication process is the best known, it is estimated that the process took more than 3000 years (Iltis, 1986). If, for an annual plant, so much time was spent, one can imagine how long it took to domesticate the Amazonian fruit trees that are perennials. The appearance of a domesticated plant may also be sudden due to some mutation that provokes a very radical modification in the plant, such as alterations in the number of chromosomes. For example, cupuaçu (*Theobroma grandiflorum* Schum) is an Amazonian fruit highly valued for its pulp, but whose seeds constitute 32% of its total weight. A triploid variety that produces only pulp and no seeds has been found and is being propagated by grafting (Venturieri and Mendonça, 1985). Peach palm (*Bactris gasipaes* H. B. K.), considered to be the only palm tree domesticated in the Americas, is probably a hybrid between *Bactris dahlgreniana* Glassman and *Guilielma insignis* (Clement, 1988). Finally, there was the suppression of some gene or group of genes that made a species poisonous, as is the case with wild manioc, with a high level of cyanidric acid, producing macaxeira with very low levels of cyanidric acid. Nowadays, with the help of advanced selection and biotechnology techniques, the process of domestication can be considerably shortened, but still 99% or more of the species currently domesticated were achieved by farmers before the "green revolution" (Harlan, 1975).

The greater the diversity of these plants, in terms of species as well as of varieties within each of these species, the greater would be the options of choices by humans.

For selection to be efficient there must be a reproductive isolation between the population selected and the original group from which it was taken. If there is no isolation, the new population again crosses with the original population and becomes similar to it, without achieving any gain. The more efficient this isolation, the more rapidly a given population may be domesti-

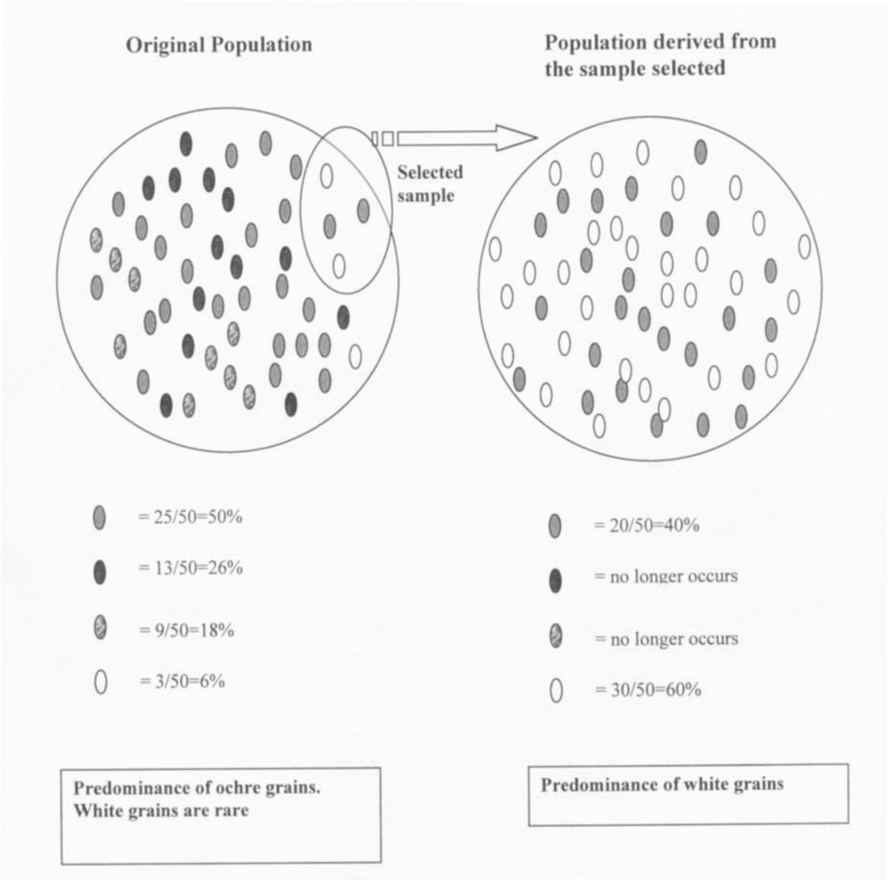


FIGURE 1 Hypothetical representation of genetic drift for the grain color character, considered not to have adaptive value. Genetic drift occurs when the sample is small and does not represent all the classes of a determined characteristic.

cated (Fig. 2). In indigenous peoples' systems, this isolation is achieved by environmental barriers or management practices. In summary, to obtain a domesticated plant there must be (i) the chance for choice (genetic diversity), (ii) discovery of useful plants and obtainment of interesting varieties and (iii) reproductive isolation of the varieties selected so that they do not return to the wild state. How were these predicates achieved for Amazonian plants?

A. The Chance for Choice (Genetic Diversity)

The Amazonian forest offers a grandiose biodiversity with an immense number of useful plants for domestication. Considering only the fruiting species (Benza, 1980; Cavalcante, 1991; Clement, 1999; Correa, 1926; Hoehne, 1946; Kerr and Clement, 1980; Kerr and Posey, 1984; Le Coint, 1947; Smith *et al.*, 1992), one may prepare a list with more than 250 typically Amazonian species, some of which are internationally recognized, others with use restricted to Indians who extract them from the native forest, and other "little-known" species in general use among the civilized population that has undergone Indian influence. Certainly, if one considers the vastness of Amazonia with its 4 million km^2 and its incredible variability of plant species, estimated to be between 50,000 and 60,000 for the higher plants alone (Clement *et al.*, 1982), and the current state of economic botany in the region, it may be concluded that this list may still be increased considerably. For example, in two expeditions specifically planned for collecting potential fruiting species in the Amazon region, 10 species were attained that had not yet been cataloged (Instituto Nacional de Pesquisas da

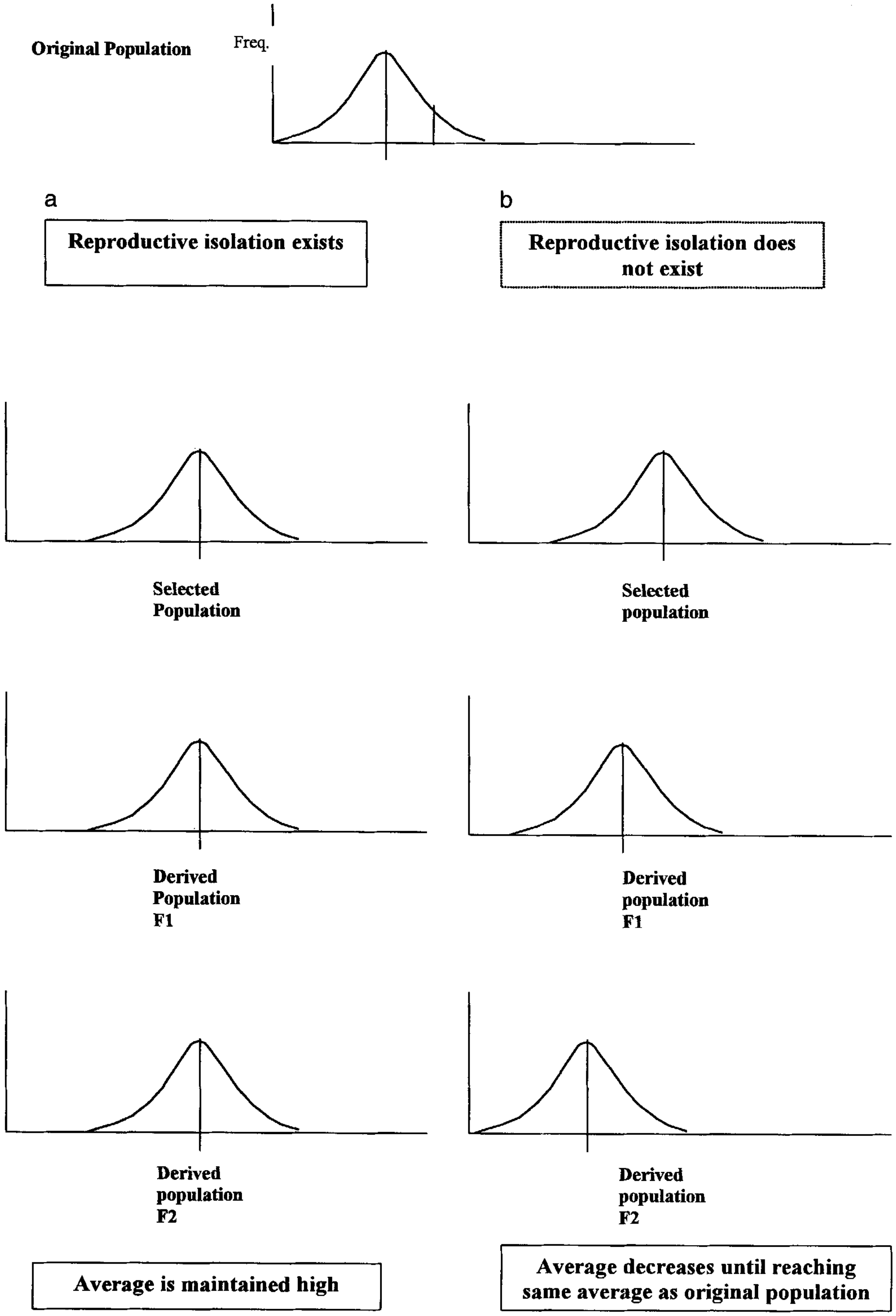


FIGURE 2 Diagram exemplifying the effect of reproductive isolation. A selected sample with large fruits is removed from an original population. (a) When the isolation is efficient—that is, when the new population is planted in a plot at a distance that avoids gene exchange with the original population—the new population will produce descendants with large fruits. (b) When isolation is not efficient, the new population may cross with individuals from the gene group that gave rise to it and reestablish the characteristics of the original population among its descendants. The more effective the reproductive isolation, the more rapidly genetic gains are achieved.

Amazonia/Núcleo do Acre, 1984; Venturieri *et al.*, 1986/1987).

B. The Discovery of Useful Plants and Obtainment of Interesting Varieties

In the eyes of "civilized" society, it is difficult to understand the process of discovering uses for plants. The Indians experiment with almost everything new that they find in the forest or that is brought to them by white people. As presents, they like receiving new plant species and varieties, such as the exotic plants that have recently been introduced in their cultures. The Kayapó Indians are very fond of mango, an Asian tree, and plant the seeds everywhere (Kerr and Posey, 1984), as do the Nambiquara Indians. Some Indians describe methodologies for discovering the usefulness of plants by employing concepts that are difficult to understand; for example, an Indian explained to me that the discovery that the papo-de-peru plant (*Aristoloqia* sp.) served as an antidote to snake venom was due to the fact that "its flower looked like the skin of a snake." It is common for them to risk their lives for discovery. I witnessed among the Hahaintesu group of the Nambiquara the ingestion of the aril of seeds of a *Philodendrum* sp. (Fig. 3), the latex and fruit of which they recognized as very poisonous and capable of causing death. They would split the fruit and with tweezers carefully remove the seeds so that they were not contaminated by the latex and then place them in their mouths to suck the arils, doing so carefully so as not to bite the seeds which were later discarded. These Indians eat tarantulas (Setz, 1991) after singing them to remove hairs that can cause irritations to the skin (in this case, in the mouth). They also cook and eat the tuberous stalk of a plant of the genus *Zamia*, which if eaten raw causes a strong headache (Fig. 4). Apparently, the process of discovering a plant's usefulness is by trial and error since the analogies between aspect and use are not logical. From the amount of information existing on edible and medicinal plants, one may imagine how many lives were lost and pains felt by these intrepid researchers.

Domesticated plants are generally described as having "appeared" through some myth. Among the Hahaintesu group of the Nambiquara there is a myth that

FIGURE 3 *Philodendrum* sp. poisonous species whose seeds are removed from the fruit with the aid of tweezers and the aril is sucked carefully so as not to bruise the seeds, which are also toxic.

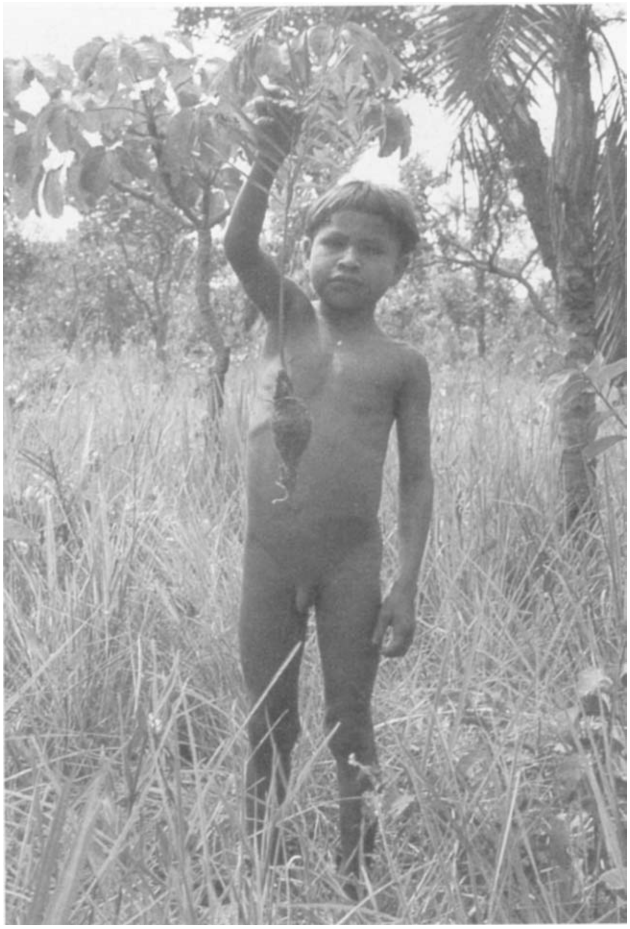


FIGURE 4 Indian boy of the Hahaintesu group of the Nambiquara showing a plant of the genus *Zamia*, whose tuberous stalk can only be eaten cooked. When ingested raw, it causes a strong headache.

describes the origin of man, whose head was the gourd (*Crescentia cujete* L.), whose teeth were the corn (*Zea mays* L.), whose ear was the lima bean (*Phaseolus lunatus* L.), whose bones were the manioc (*Manihot esculenta* Crantz), and whose scrotum was the cará yam (*Dioscorea trifida* L.) (Fig. 5). In the Kayapó myth, corn was acquired from the stomach of a rat that lived in an enormous tree in the middle of an area that was being prepared for planting; the strength of the entire tribe was needed to fell the tree (Lukesch, 1976). Another Kayapó myth mentions that sweet potato, macaxeira, and banana were brought by the "daughter of the rain" who came down from the sky, became an ordinary Indian, married, had children, and during a time of

FIGURE 5 Species from the Nambiquara myth about the appearance of cultivated plants cultivadas (a) the head is represented by the calabash (*Crescentia cujete* L.) (photo by Nigel J. H. Smith), (b) the teeth by corn (*Zea mays* L.), (c) the ear by the lima bean (*Phaseolus lunatus* L.), (d) the bones by manioc (*Manihot esculenta* Crantz), and (e) the scrotum by the *cará* yam (*Dioscorea trifida* L.).

hunger returned to the sky and brought back these species that were then cultured so that the tribe never again went hungry (Lukesch, 1976). Others describe useful plants as being inherited from a conquered people. The Tikuna recognize that many of the obviously domesticated fruits were already present when they arrived, possibly derived from the Omagua Indians, whom they expelled (Ailton Krenak, União das Nações Indígenas, personal communication), or they negotiated with other tribes (among the Hahaintesu group of the Nambiquara there is a variety of manioc called the "manioc of the Wassuso," a tribe from the same linguistic group that is practically extinct: in 1985, there was only one elderly woman of this tribe).

The Indians' diet is based on many plants and their varieties. One commonly finds up to 40 varieties of manioc per indigenous tribe (Kerr and Clement, 1980; Boster, 1984) and witnesses the dynamism these peoples have for renewing these varieties. Manioc is usually propagated by stakes (cloned) which avoids the combination between genotypes. However, if this were always they case, how would recruitment of new varieties for selection exist? How would one justify the great number of varieties observed among the Indians? Contrary to what is found in the literature, manioc is also propagated by seeds (Kerr and Posey, 1984). Plantules derived from seeds bring possibly new combinations resulting from crossing parents propagated by stakes. The Indians of the upper Rio Negro recognize plantules of manioc as derived from seed banks of old plantings in areas that have regenerated and been slashed and burned anew. This recognition is easy because the plantules derived from seeds display their cotyledons (Paulo Sodero Martins, Escola Superior de Agricultura "Luíz de Queirós," personal communication).

The Hahaintesu group of the Nambiquara in 1985 possessed only nine varieties of macaxeira, almost all recently introduced—after their areas were demarcated and their extermination by whites was interrupted. The increase in varieties is an indication that peace may be giving the Indians a chance to recompose their genetic collection of plants (Boster, 1984). Generally, as in the case of birth control, the administration of genetic resources is directed by women (Kerr and Posey, 1984). In the case of the Hahaintesu group of the Nambiquara, an elderly Indian woman (as mentioned earlier, the last of the Wassuso, married to the Indian Papai) possessed a great knowledge and love for agricultural plants, and others frequently brought her new plants and varieties. This woman would plant them in different environments and evaluate them. Those that were approved she handed over to the rest of the tribe for planting. This Indian woman and her garden fulfilled what would be the role of "curatorship and experimental station for genetic resources" in our society.

The Amazonian Indians highly value aesthetics, as expressed through their clothing, body decorations, and basketry, which are generally exuberant in both form and color. This sentiment is also reflected in the plants and varieties domesticated by them. Examples include peach palm (*Bactris gasipaes* Kunth) (Fig. 6A) and peppers of the genus *Capsicum* (Fig. 6C) and cubiu (*Solanum sessiliflorum* Dunal) (Fig. 6E), which have a large variety of colors, sizes, and shapes. Among Indians, selection is also made by productivity and taste. Peach palm, one of the best domesticated plants of the humid tropics, can produce up to 50 tons per hectare, with varieties selected for flour or oil and others that easily release the skin and have a refined flavor for cooking; there are also varieties with and without spiny trunks (Clement, 1988; Clement and Mora Urpi, 1987). Cubiu (S. sessiliflorum Dunal), a fruiting plant similar to tomato, produces up to 100 tons per hectare and has varieties with eight different shapes ranging from fusiform to semi-rounded, with colors ranging from yellow to purple (Silva Filho *et al.*, 1993, 1999).

New varieties are also achieved through barter, raiding, and dower. Groups of friendly Indians visiting from other tribes take varieties as presents. Rival Nambiquara trade varieties in a ritual that is difficult to understand. They peep at each other from opposite sides of a river, without exposing themselves so as not to be killed; they sing and call out challenges. Suddenly, they gain confidence and cross the river to leave presents that sometimes are repaid. After the exchanges they go back to being mortal enemies. Among the Alantesu group of the Nambiquara, after marriage is consented to for a woman, some compensation is required from the groom, which may take the form of labor, a bicycle, a radio, or some other article or group of articles. When the debt is paid, the husband takes his wife and with her some plants, which introgresses genotypes from other tribes. Kidnapping is very common among the indigenous tribes of Amazonia (Black *et al.*, 1991). They kidnap children (the chief of the Ureu-au-au Indians, shown in a 1970s documentary about colonization in Amazonia along the BR-364 highway in the state of Rondônia, was a white man without any cultural traces of white civilization). Women are kidnapped, which seems to be a common practice among the Indians that does not annoy the kidnap victim for very long. Among the Hahaintesu group of the Nambiquara, we saw a kidnapped woman who was imprisoned in a straw hut from which escape would have been easy and was

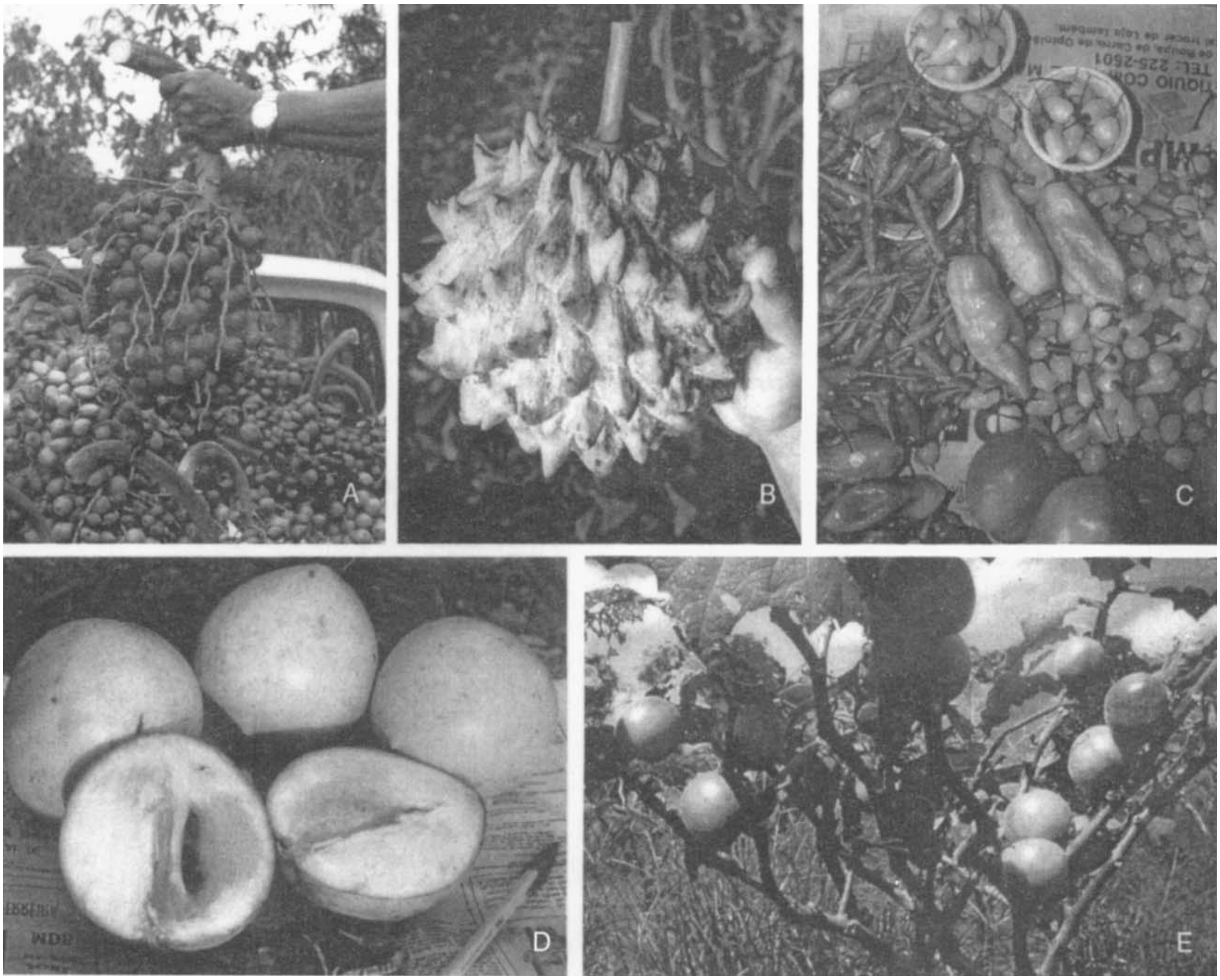


FIGURE 6 Good examples of the great variety of colors, sizes, and shapes of domesticated plants selected by Amazonian Indians: (A) Peach palm (*Bactris gasipaes* Kunth), (B) *Biribá* (*Rollinea mucosa* Bail.) var. Tikuna (photo by Charles R. Clement), (C) peppers of the genus *Capsicum;* (D) *abiu* (*Pouteria caimito* Radl.) var. Tikuna (photo by Charles R. Clement), and (E) *cubiu* (*Solanum sessiliforum* Dunal). See also color insert, this volume.

guarded by her kidnapper, who stayed nearby. During the first days, she was aggressive toward those who approached her; two days later, we saw her laughing among the other Indian women. I later learned that she returned to the agricultural plots of her original tribe and brought back new varieties of manioc, and thus increased her standing with her new tribe.

Domesticated plants return to their wild state if the selective human pressure is relaxed (Harlan, 1975); therefore, there is a close relation between humans and domesticated plants. This relation is so intimate that gene introgression observed in Indian populations in Amazonia (Black *et al.*, 1991) may be associated with the introgression also found among populations of domesticated plants, such as is observed in peach palm (*B. gasipaes* Kunth) (Picanço *et al.*, 1999) and in caiaué (Elaeis oleifera Cortés) (E. Barcelos, Embrapa Amazonia Ocidental/Manaus, personal communication, and data partially published in Barcelos *et al.*, 1999).

Among the Tikuna in the region near Leticia, Colombia, there is a tribal custom that reinforces plant selection: "All fruit or other edible product that is especially large or flavorful is divided among the tribe and seed planted." This process practiced over many generations led, by massal selection, to an increase in desirable alleles, giving rise to the domestication of fruiting species such as peach palm (B. gasipaes Kunth; Fig. 6a), biribá (*Rollinea mucosa* Bail.; Fig. 6b), and abiu (*Pouteria* *caimito* Radl.; Fig. 6D) (Kerr and Clement, 1980). The fruiting species of that region have very beautiful varieties with giant fruits, showing clear signs of being domesticated; therefore, this is suggested as a center of domestication (Clement, 1989).

III. THE SELECTION, FIXATION, AND STATUS OF AMAZONIAN PLANT DOMESTICATION

A. Reproductive Isolation

For one population to be different from another, it must fragment off from the original and in some manner maintain itself distinct in genetic terms: in other words, it should have very low or nonexistent chances of crossing with individuals of the population from which it originated. Any reproductive barrier is capable of permitting the fragmentation of a population and its differentiation into sub-populations. Many isolation mechanism are known from cultivated plants: geographic and ecological separation, differences in time of blooming, self-fertilization, translocation races, polyploid races, gametophytic factors, cryptic chromosomal differences, and meiotic irregularities (Harlan, 1975). These mechanisms may occur simultaneously and be perfect, and if the sub-populations meet someday some degree of hybridization may occur. Sub-populations with potential for intercrossing are called varieties. When the isolation is perfect and there are no longer possibilities for a natural cross to occur, then with the generation of fertile progenies a new species has been formed.

B. The Environment

Practices capable of inducing modifications in the structure of a population are intimately related to the physiognomy of the site and to the ecological restrictions that each environment offers. In Amazonia, although there is a range of ecosystems (for climate, see Salati, 1985, for soils, see Jordan, 1985; and for vegetation, see Pires and Prance, 1985) and a varied number of singular strategies used by the Indians who practice agriculture on lands with very low trophic support—such as the use of barreais near the rivers (Hill and Moran, 1983) and in the savannas, with the formation of "managed islands" or apêts of the Kayapó (Anderson and Posey, 1989)—agriculture in indigenous communities is basically practiced on (i) upland areas (terra firme) covered by forests and (ii) the river floodplains (várzeas).

On terra firme covered by forest, agriculture is practiced in cycles of felling–burning–planting–fallowing, also called slash-and-burn or swidden fallows. This is the most applied system of agriculture in Amazonia. A site with exuberant forest is located and the understory trees are cut first, followed by the large ones. Three months later, fire is applied. The ashes and organic matter formed by decomposing tree trunks nourish the soil for at most 4 years. When these soils are exhausted, cultivation of annual plants is abandoned and a new cycle must begin. Among Indians and many caboclo communities, this is not a complete abandonment because there is a subsequent recuperation stage of planting with useful perennial plants (Denevan *et al.*, 1984). This system of apparently migratory agriculture has been very efficient in improving plants because each plot receives some seeds of selected plants. Because the plots are relatively isolated from each other as if islands within the forests and the majority of plants cultivated in them do not occur inside the forest, the latter serves as a barrier to gene flow (Clement, 1987) (Fig. 1). On terra firme, the soils are generally poor and need fertilizing, generally with the ashes resulting from the felling and burning of the original forest that covered them. On terra firme there are also areas that are cultivated continuously, although they are very rare. These are the *terra-preta-do-índio* (Indian black earth) anthropomorphic soils that are very fertile due to having been enriched by organic detritus deposited over countless generations by the Indians who inhabited them (Smith, 1980) (Fig. 7). These areas vary from 0.5 to 5 ha in area and due to their fertility are preferred even today by white people for planting. Along the Andirá and Paraná-do-Ramos rivers, in the municipality of Barreirinha, AM, Brazil, sites with Indian black earth area are abundant. Some of them can be found in the municipal seat and in the communities of Ariaú, Freguesia, Maloca, Terra-Preta-do-Limão, Terra-Preta-do-Castanhal, and Pedras Atalaia (Francisco das Chagas Rodrigues Júnior and João Carlos dos Anjos Pimentel, residents of Barreirinha, personal communication and author's observations).

Agriculture on periodically flooded lands is practiced on high and low várzeas and on riverbanks in cycles dictated by seasons of high and low waters. Despite this inconvenience, these lands are highly fertile and demand little labor for clearing and planting. Annual plants and a plantain-type of banana (*Musa paradisica* L., possibly the variety known as *pacovi*) with small bunches and an extremely short productive cycle are

FIGURE 7 Anthropogenic soil of the "Indian black earth" (*terra-preta-do indio*). Its great fertility is caused by the accumulation of ashes, wastes, and excrement of domestic animals. Many perennial and annual plants are cultivated on these soils. Andirás river, Barreirinha, AM, Brasil.

cultivated (Cavalcante, 1991). On the high várzeas, where flooding rarely occurs, perennial plants that can survive some time with roots submerged and annual plants are cultivated. With some risk, perennial plants typical of terra firme are also cultivated on high várzeas (Smith, 1996).

These two groups of ecosystems have caused a great division between groups of species and varieties of domesticated plants. For the plants most domesticated among Indians, such as manioc (*M. esculenta* Crantz), corn (*Z. mays* L.), and lima bean (*P. lunatus* L.), short-cycle varieties were developed capable of being cultivated on the várzeas before the floods. Although plantain banana (*M. paradisica* L.) is an exotic and perennial species, the Indians have also selected varieties and cultivation systems of it that are adapted to the várzea. For plants propagated by stakes, such as manioc (*M. esculenta* Crantz), when floods arrive the young plants are removed from the soil and placed in a canoe or higher site for planting when the waters recede. In the case of banana plants, marombas (wooden structures elevated above the water level) are made in which the young plants are placed with a little soil and fertilizer so that they are sufficiently grown to be planted when the waters recede. There are also the *tesos* of Marajó Island, which are hillocks made by the Indians in periodically flooded areas, where possibly corn was cultivated (Roosevelt, 1991). There is insufficient information about how these tesos were cultivated and what species were utilized, but presumably there must have been a great food production capacity capable of having maintained a quite dense population. It is estimated that at the time of discovery, on terra firme there were 0.2 inhabitants per square kilometer and on the várzea 14.6 (Denevan, 1976)—numbers that are higher than those currently observed in Amazonia.

C. The Management System

The management system causes environmental changes that can influence the demography and genetic structure of a population. For there to be selection and consequent domestication, there must be differentiated reproduction and survival for the genotypes selected; therefore, management and domestication are two intimately linked factors. Forms of management generally influence domestication and it is through them that I intend to explain this process among Amazonian plants. Generally, the more elaborate the management demands of a plant, the greater its degree of domestication. In the case of Amazonian plants, depending on the degree of management to which they may be submitted, they may be classified as (i) collected, (ii) protected or planted with little human interference, and (iii) planted with high human interference. There is also another distinct group—those managed unconsciously due to interference in the environment by humans with other purposes or due to dispersion of seeds through feces and/or discarding of non-edible parts.

Plants that are simply collected are left in their natural habitat without selection for any characteristic, and even if the fruits collected from more productive and better plants should be more disseminated there is no reproductive isolation and presumably the population maintains itself stable. The word "presumably" is used because no matter how subtle, human occupation is capable of introducing some modification in the natural populations of plants. The reduction, through hunting, of frugivorous animals such as agoutis (*Dasyprocta leporina* L., *D. prymnolopha* Wagler, and *D. fuliginosa* Wagler), pacas (*Agouti paca* L.), and bats, which prey on and disperse seeds, may favor the increase in density of a certain species where the more productive plants might have greater chances of increasing the frequency of their alleles in a given location due to the increase in their density. The group of useful plants collected in the forest for medicinal purposes, food, or some other use may possibly comprise more than 1000 species.

The plants protected or planted with low human interference are those that occur naturally in the habitat and that are preserved during felling and later cared for or planted along trails or in clearings but that receive little care (Table I). Examples of plants protected from burning include the tucumã (*Astrocarium aculeatum* G. F. W. Meyer and *Astrocarium vulgare* Mart.) and inajá [*Maximiliana maripa* (C. Serra) Drude] palms. Plants under low human interference are those generally cast along trails and in clearings, such as is done by the Kayapó Indians, who collect seeds and plantules of useful trees and plant them along trails on which they usually travel or in clearings opened by tree falls (Kerr and Posey, 1984). They generally belong to the same habitat and although they undergo some selection, reproductive isolation, as in the previous case, is not efficient because these species also occur in forests and can interbreed. Occasional progress in the domestication process may be achieved because the selected plants, when planted together, have greater chances of intercrossing, which facilitates the maintenance of a higher frequency of the alleles selected. Since isolation is not very efficient, progress is very slow. This group is prone to receive greater human influence than is the case with the previous group because their density and fecundity are in most cases favored by the reduction of competition with the other plants. The reproductive isolation of these species, although more efficient than in the previous group, is also not very efficient and there is no directional selection of the plants; they are simply spared.

The dispersion of those planted with high human interference (Table II) is consciously done by humans and during their entire productive cycle they receive care and make up three distinct groups, according to the position they assume in the ecological succession of forest regeneration after burning. This succession is directed by humans toward a final floristic composition with a greater density of useful species than would be found in a forest without any history of human influence (Denevan *et al.*, 1984; Anderson and Posey, 1989; Balée, 1989). The first groups is that of the most domesticated species: These are annual plants that are colonizers and fast growers, with a high productivity of edible biomass. They are: manioc (*M. esculenta Crantz*), corn (*Z. mayz L.*), and more recently, beans of the genus vigna. The second group is formed by semi-perennials species, sown after the plants in group 1: sweet potato (*Ipomoea batatas* L.) Lam., arrowroot (*Marantha arundinaceae* L.), peanut (*Arachis hypogeae* L.), lima bean (*P. lunatus* L.), cará yam (*Dioscorea trifida* L.), and pineapple [*Anannas comosus* (L.) Merril]. Those of the third group are arboreal species, generally fruit producers with rapid growth and the capacity for growing well in soils already leached by rains and exhausted by annual crops. These plants have a high capacity for production and deposition of their leaf biomass. They live for 80 years at the most and cannot bear competition with other trees of the forest. They are peach palm (*Bactris gasipaes* H. B. K.), cucura (*Pourouma cecropiaefolia* Mart.), umari do Amazonas (*Poraqueiba sericea* Tul.), and biribá *Rollinia mucosa* (Jacq.) Baill. Also in this group are other trees that do not have a great capacity for regenerating the soil but are still able to grow well in poor soils: guava, coca, and abiu. After the planting of these categories, there is the reappearance of the group of plants protected/planted under low human interference. These make up the group in the final phase of regeneration in the direction of a climax forest. Sometimes, species of this group are planted, whereas others are only protected and require more specific ecological conditions, with some shading in the juvenile phase, humidity, and protection from heat. These plants are capable of competing with other trees of the forest and are part of the final climax stand achieved by the system: They are Brazilnut (*Bertholletia excelsa* H & B) and the piquiazeiros (*Caryocar villosum* Pers.), amapá doce (*Brosimum potabile* Ducke), maçaranduba [*Manilkara*

TABLE I

Protected or Planted Plants at Low Levels of Human Care[a]

Names: English/Portuguese/Spanish	Latin name	Uses
—/caruru azedo/—	*Amaranthus cruentus* L.	Leaves as vegetable
Star nut palm/tucumã do Amazonas, tucumã-açu/chambira	*Astrocarium aculeatum* G. F. W. Meyer	Fruit, fibers from leaves
—/tucumã-do-pará/aouará, awarra	*Astrocarium vulgare* Mart.	Fruit; fibers from leaves
Brazil nut/castanha-do-pará/nuez del Brasil	*Bertholletia excelsa* Humb. Et Bonpl.	Nuts; infusion of fruits as a medicinal
Wild cherry/muruci verdadeiro/Indano, nanci, yoco	*Byrsonima crassifolia* L. Rich.	Fruit
Pekea/piquiá verdadeiro/amêndoa-do peru, almendro	*Caryocar villosum* (Aubl.) Pers.	Fruit, and oil from the pulp
—/castanha-de-porco/—	*Caryodendron amazonicum* Ducke	Seed as a nut
—/—/—	*Celosia argentea* L.	Leaves as vegetable
—/castanha-de-cutia/—	*Coepia edulis* Prance	Fruit
—/pajurá/—	*Couepia bracteosa* Benth.	Seeds as a nut
—/castanah pêndula/—	*Couepia longipendula* Pilger	Seeds as a nut
—/sorva, sorvinha/—	*Couma utilis* Muell. Argov.	Fruit
American oil palm/caiaué/noli, palma brasilenha, carocito, colorada, palmiche	*Elaeis oleifera* Cortés	Oil from the pulp and endosperm, seed kernel for crafts
—/uxi/—	*Endopleura uchi* Cuatr.	Fruit
Wild coriander/chicória, coentro do pará/siuca culantro	*Eryngium foetidum* L.	Condiment
Assai palm/açaí/chonta, husaí	*Eutherpe oleracea* Mart.	Fruit, trunks for building
—/mangaba/mango jsú	*Hancornia speciosa* Gomes	Fruit
Copal/jatobá/copalhuallo	*Hymenaea courbaril* L.	Fruit
—/ingá-açu/—	*Inga cinnamomea* Benth.	Fruit
Ice cream beam/ingá-cipó/guamo, guabo	*Inga edulis* Mart.	Fruit
Pataua/patauá/bataua, chapil, palma sege	*Jessenia bataua* (Mart.) Burr.	Fruit
Barbasco/timbó/barbasco	*Lonchocarpus utilis* Smith	Poison for fishes
—/maçaranduba, maçaranduba verdadeira/—	*Manilkara huberi* (Huber) Standl.	Fruit, edible latex
Mauritia, moriche/buriti, miriti/aguaje	*Mauritia flexuosa* L.	Fruit, sap as sweet syrup, starch from the trunk

continues

huberi (*Huber*) Standl.], bacuri (*Platonia insignis* Mart.), cacao (*Theobroma cacao* L.), cacauí (*Theobroma speciosum* Willd.), and uxi (*Endopleura uchi* Cuatr.). Because trees of the climax forest and those of this group have the chance to make gene exchanges, diluting eventual gains in selection, they cannot be considered domesticated. It is curious to note that young Brazil-nut trees are only seen in disturbed environments, but that in the climax forests they are frequent, which reinforces the hypothesis that the Brazil-nut forests, such as those observed in the regions of southern Pará, Acre, and Alenquer, had once been various Indian agricultural plots. These useful forest trees form the so-called "maçaranduba islands," "Brazil-nut islands," "amapá–doce islands," "bacuri islands," "cacao islands," "cacauí islands," "uxi islands," and so on. When they begin to produce fruits, these islands begin to receive better care. The plants below the canopy are weeded to facilitate fruit collection and observation of game that also come searching for fruits. The Nambiquara Indians of the Hahaintesu have a curious system for transferring information on the localization of these "islands" from one generation to the next until they are ready to be exploited. The sites where they live are eventually abandoned, generally due to over-hunting or disease. The new site chosen is generally an old site that had been used because it had a water supply, offered some strategic defense position, and had trails enriched with useful plants. The former site has a cemetery. The tribe is divided into peoples: the "people of the curassow," "the

Continued

Names: English/Portuguese/Spanish	Latin name	Uses
—/bacaba açu, bacaba verdadeira/ungurauy, camou, manoco	*Oenocarpus bacaba* Mart.	Fruit, truck for building, leaves to cover houses
—/bacaba-de-leque/—	*Oenocarpus disticus* Mart.	Fruit, truck for building, leaves to cover houses
—/bacabi/mapora	*Oenocarpus mapora* Karsten	Fruit
—/bacabinha/—	*Oenocarpus minor* Mart.	Fruit
—/feijão macuco/—	*Pachyrrhizum tuberosum* Lam.	Tubers as vegetable
—/pajurá-da-mata/—	*Parinari montana* Aubl.	Fruit
Wild passion fruit/maracujá suspiro, maracujá-de-rato, maracujá-do-mato/—	*Passiflora nitida* H. B. K.	Fruit
—/camapu/—	*Physalis angulata* L.	Fruit as a vegetable, roots as medicine
Backuri/bacuri/bacuri, matazona	*Platonia insignis* Mart.	Fruit
—/umari, mari gordo/—	*Poraqueiba paraensis* Ducke	Fruit
—/umari, umari do amazonas/—	*Poraqueiba sericea* Tul.	Fruit
Canistel/cutite/caimo, canistel, yema de huevo	*Pouteria macrophylla* (Lam.) Eyma	Fruit
—/ucuqui/—	*Pouteria ucuqui* Pires et Schultes	Fruit
—/araça-pera/—	*Psidium acutangulum* DC	Fruit
Guavas/goiaba/guayaba	*Psidium guajava* L.	Fruit, medicinal tea from young leaves for diarrhea
—/araçaí, araça-do-campo/—	*Psidium guinensis* Swartz	Fruit
—/bacuripari liso/—	*Rheedia brasiliensis* (Mart.) Pl. et Tr.	Fruit
—/bacuripari/pacuriguazu	*Rheedia macrophylla* Pl. et Tr.	Fruit
Hog plum, plum bush/taperebá, cajamirim, cajá/mombim	*Spondias mombim* L.	Fruit, infusion from the bark for leucorrohea
—/pitomba/carayá-vola	*Talisia esculenta* (St. Hil.) Radlk.	Fruit
Cocoa, chocolate tree/cacau/cacao	*Theobroma cacao* L.	Fruit
Cupuassu/cupuaçu, cupu, cupuácu verdadeiro/copoazur	*Theobroma grandiflorum* (Willd. ex Spreng.) Schum	Fruit
—/cacauí/cacau sacha, chocalatillo	*Theobroma speciosum* Willd.	Fruit

[a] They receive some influence on the genetic structure of their population. The domestication process is accepted as being in progress but at a very early stage. Some of these species were possibly undergoing a more advanced process of domestication, but a possible retrocession on their status is accepted due to loss of varieties or by the people who used them (table structured from Denevan *et al.*, 1984; Clement, 1999; Cavalcante, 1976; Smith, 1999; Smith *et al.*, 1992).

jaguar," "the tapir," etc., and the individuals of each "people" have a specific cemetery where they should be buried. When Indians die, their companions take them to be buried in the cemetery of their respective people and thus they are able to visit the old sites, accompany the development of the "islands," and maintain their localization in tribal memory.

The groves of babaçu palms (*Orbignya martiana* Barb. Rod.), which occupy approximately 200,000 km^2 in Brazil and support approximately 450,000 subsistence-level households in the north, northeast, and central west regions of Brazil through their multiple products (fibers, thatch, basketry, construction, palm heart and fruits for oil, charcoal, and animal feed) are an example of unconscious management. This palm tree regenerates profusely when the primary forest is burned. The plant survives fire well because it has a unique germination process. Upon germinating, the seed sends its hypocotyl underground to a depth of about 10 cm and then growth of the gemule begins. It remains inside the soil, putting its leaves out aboveground. When the forest is burned, the plants that are still in this phase lose only their leaves but are capable of regenerating since the main bud is well protected from the fire. Due to the reduction in competition with the other plants that were destroyed by fire, the babaçu plants end up dominating the environment (May *et al.*, 1985a, b). Another example of unconscious

TABLE II

Cultivated Plants under a High Level of Human Assistance[a]

Names: English/Portuguese/Spanish	Latin name	Uses
Cashew/caju/marañón	*Anacardium occidentale* L.	Fruit, infusion from the bark for leucorrohea
Pineapple/abacaxi, ananas/piña	*Anannas comosus* (L.) Merril.	Fruit, fiber from leaves
Soursop/graviola/guanábana, anona	*Annona muricata* L.	Fruit, infusion from the leaves for snake bite
Peanut/amendoin/mani	*Arachis hypogeae* L.	Seeds
Peach palm/pupunha/pejibaye, pijuayo, chontaduro	*Bactris gasipaes* H. B. K.	Fruit, trunks for building
Annatto/urucum/achiote	*Bixa orellana* L.	Arielle from seeds as colorant and condiment, largely used for painting the body
Guinea arrow root/ariá/dale dale	*Calathea allouia* (Aubl.) Lindl.	Tubercle
—/—/—	*Canna edulis* Ker.	Tubercle
Sweet pepper/pimenta-doce/ají	*Capsicum anumm* L. var. minimum	Fruit as a vegetable, condiment, medicinal
—/—/—	*Capsicum chinense* Jacq.	Condiment
—/pimenta-de-cheiro/—	*C. brasilianum*	Condiment
—/pimenta comari/—	*C. baccatum* L.	Condiment
Papaya, paw paw/mamão/lechosa, melon zapote	*Carica papaya* L.	Fruit
—/cupa/—	*Cissus gongyloides* Burch	Liana as vegetable
Calabash gourd/cuieira/pati	*Crescentia cujete* L.	Fruit domestic appliances
Yam/cará/sacha-papa	*Dioscorea trifida* L.	Tubercle
—/jaboti/—	*Duguetia stenantha* R. E. Fries	Fruit
—/araça-boi/arazá, araza buey	*Eugenia stipitata* McVaugh	Fruit
Genipa/genipapo/huito	*Genipara americana* L.	Fruit
Cotton/algodão brabo/algodón	*Gossipium barbadense* L.	Fiber
Sweet potato/batata doce/camote	*Ipomoea batatas* (L.) Lam.	Tubercle
Manioc/mandioca/yuca	*Manihot esclenta* Crantz	Tubercle, leaves as vegetable
Arrowroot/araruta/—	*Marantha arundinaceae* L.	Tubercle
Tobacco/tabaco/tabaco	*Nicotiana tabacum* L.	Leaves that are smoked or smashed as stimulant
Passion fruit/maracujá, maracujá, verdadeiro/granadilla	*Passiflora edulis* Sims. f. flavicarpa Deg.	Fruit for a refrigerant and as a carminative
Giant granadilla/maracujá-acú, maracujá mamão/granadilla, granadilla real	*Passiflora quadrangularis* L.	Fruit
Guarana/guaraná/cupana	*Paullinea cupana* H. B. K. var sorbilis (Mart.) Ducke and var. typica	Seeds for infusion used as refrigerant and stimulant
Lima beam/feijão lima/—	*Phaseolus lunatus* L.	Beans
—/cucura, mapati/uvilla	*Pourouma cecropiaefolia* Mart.	Fruit
—/abiu/caimito	*Pouteria caimito* (Ruiz et Parvon) Radlk.	Fruit
Sapote/sapota-do-peru/zapote, mame colorado	*Quararibea cordata* (H. B. K.) Vischer	Fruit
—/biribá/—	*Rollinia mucosa* (Jacq.) Baill.	Fruit
—/cubiu/cocona	*Solanum sessiliflorum* Dunal.	Fruit as a vegetable
Para cress/jambu/jambu	*Spilanthes acmella* (L.) Murr.	Leaves as vegetable
—/cacau-do-peru/macambo	*Theobroma bicolor* Humb. & Bompl.	Fruit, toasted seeds as a nut
Cocoyam/taioba/uncucha, malanga	*Xanthosoma sagittifolium* (L.) Schott.	Tubers and leaves as a vegetable
Corn, maize/milho/mayz	*Z. mayz* L.	Seed as a grain, estiles as medicinal

[a] The domestication process is accepted as being in an advanced stage (table structured from Denevan *et al.*, 1984; Clement, 1999; Cavalcante, 1976; Smith, 1999; Smith *et al.*, 1992.)

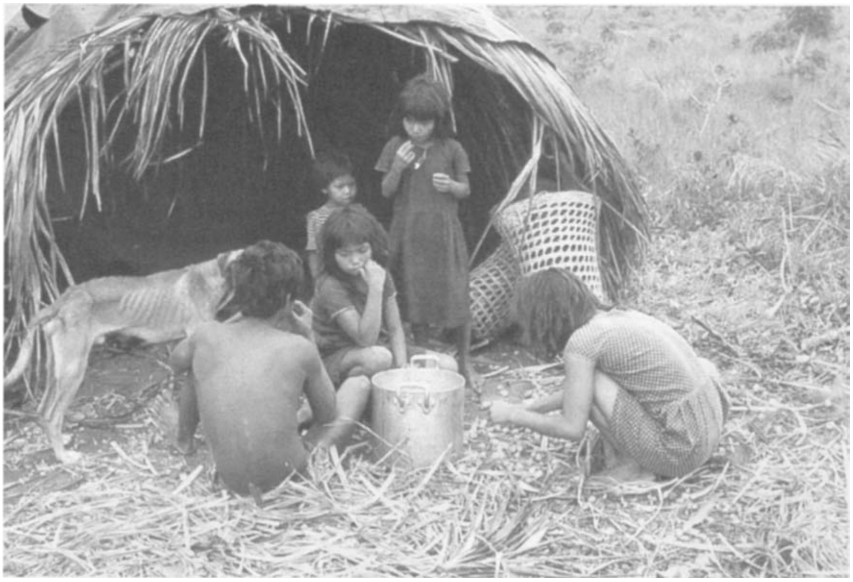


FIGURE 8 Nambiquara Indians of the Alantesu group enjoying fruits recently collected from a fruiting plant in the savanna and scattering the seeds near a temporary hut built for support on one of their collecting journeys.

management is the sowing of plants collected near the huts in piles of trash (Fig. 8). This process is possibly what happened at the beginning of agriculture and the related sedentarism (Harlan, 1975). The plants thus unconsciously cultivated end up being cared for and recruited within an agricultural system.

All these system of management and consequent selection have been applied since before the discovery to the present by the indigenous peoples, later by the caboclos, and recently by Japanese immigrants to Amazonia (Saragoussi *et al.*, 1988). This has maintained the preservation of already domesticated plants and their varieties and even the appearance or improvement of others. The Japanese farmers of Tomé-Acú and Castanhal and their descendants, using only massal selection, have been able to achieve better cultivars of cupuaçu (*Theobroma grandiflorum* Schum.) and açaí (*Euterpe oleracea* Mart.). Virola softwood from the estuary islands of the Amazon River, which in the past was predated as a source of extractivism, today is wisely managed as a consequence of market forces, and thus is also undergoing domestication by caboclos.

IV. FINAL COMMENTS

A domesticated plant, it should be remembered, is in this state because humans brought about its domestication. If human-induced selection pressures cease, these plants will not survive, or their descendants will need to recompose their defenses to face competition in a natural environment and in doing so generally lose the attributes selected by humans (Harlan, 1975). Many of the indigenous cultivars were selected not only for flavor or productivity but also for mystical reasons and are thus susceptible to disappearance as a result of changes in habits. In Amazonia, the Indians are being besieged by pressures toward deculturation from the "civilized" peoples who are occupying the region. In this process much of the existing knowledge on plant use is lost. For medicinal plants, this loss is even greater since most of the time their usefulness is not visible, as in the case of fruiting species. With the gradual loss by the Indians of the properties and uses of the plants, everything may simply be considered "jungle." Although Amazonia is considered one of the largest repositories of genetic resources in the world, there are still few efforts in recovery, collection, and evaluation of useful plants among the Indians. If humanity has these marvelous plants, delights in their flavor and their appearance, wears clothes and uses domestic utensils made from them, and uses them to cure pains and diseases, this is because the traditional peoples of Amazonia, very simple people, discovered them, selected them, and maintained them until now. The society that calls itself civilized still has not sufficiently recognized this feat and its value, but it must some day find mechanisms for doing so or will regret its lack of action in the same way as we today regret the loss of the Alexandria library, burned down in the name of a culture that judged itself the owner of superior values.

Acknowledgments

I am enormously grateful to Indian agent Sylbeny de Almeida and Nurse Maria Aurora (Lia) who kindly accompanied my group on our travels among the Nambiquara Indians and translated for us their customs and information about plants and myths. I also thank all the Indians and Amazon caboclos for having shared with us a little of their wonderful knowledge about their useful plants.

See Also the Following Articles

AGRICULTURE, TRADITIONAL • BREEDING OF PLANTS • DOMESTICATION OF CROP PLANTS • EDIBLE PLANTS • GENETIC DIVERSITY • INDIGENOUS PEOPLES, BIODIVERSITY AND

Bibliography

Anderson, A., and Posey, D. A. (1989). Management of tropical scrub savanna by the Gorotire Kayapó of Brazil. *Adv. Econ. Bot.* 7, 159–173.

Balée, W. (1989). The culture of Amazonian forest. *Adv. Econ. Bot.* 7, 1–21.

Barcelos, E., Cunha, R. N. V. da, Nouy, B., and Sousa, N. R. (1999). Uso de marcadores moleculares RFLP e AFLP no estudo da diversidade genética do dendezeiro (*Elaeis guineesis* Jacq. E *E. oleifera* (Kunth) Cortés). *Genet. Mol. Biol.* 22(3), 289.

Benza, J. C. (1980). *143 Frutales Nativos*. Univ. Agraria La Molina, Lima.

Black, F. L., Pandey, J. P., and Santos, S. E. B. (1991). Evidências baseadas em HLA e IgG sobre as relações intra e intercontinentais das populações nativas da Amazônia. In *Origens, Adaptações e Diversidade Biológica do Homem Nativo da Amazônia* (W. A. Neves, Ed.), pp. 55–83. Museu Paraense Emílio Goeldi (coleção Emilie Snethlage), Belém.

Boster, J. S. (1984). Classification, cultivation, and selection of Aguaruna cultivars of *Manihot esculenta* (Euphorbiaceae). In *Ethnobotany in the Neotropics* (G. T. Prance and J. A. Kallunki, Eds.). pp. 34–47. New York Botanical Gardens, New York.

Cavalcante, P. B. (1991). *Frutas Comestíveis da Amazônia*, 5th ed. CEJUP/Museu Paraense Emílio Goeldi, Belém.

Clement, C. R. (1987). A pupunha, uma árvore domesticada. *Ciência Hoje* 5(29), 42–49. (The pejibaye, a domesticated tree. *Ciência Hoje* (Special Ed.: Amazonia), 43–47, 1991

Clement, C. R. (1988). Domestication of pejibaye (*Bactris gasipaes* H. B. K.): Past and present. *Adv. Econ. Bot.* 6, 155–174.

Clement, C. R. (1989). A center of crop genetic diversity in western Amazonia. *BioScience* 39(9), 624–631.

Clement, C. R. (1992a). Domesticated palms. *Principes* 36(2), 70–78.

Clement, C. R. (1992b). Frutas da Amazônia. *Ciência Hoje* 14(83), 28–37.

Clement, C. R. (1999). 1492 and the loss of Amazonian crop genetic resources. I. The relation between domestication and human population decline. *Econ. Bot.* 53(2), 188–202.

Clement, C. R., and Mora Urpi, J. (1987). The pejibaye (Bactris gasipaes H. B. K., Arecaceae): Multi-used potential for lowland humid tropics. *J. Econ. Bot.* 41, 302–311.

Clement, C. R., Müller, C. H., and Flores, W. C. (1982). Recursos genéticos de espécies frutíferas nativas da amazônia brasileira. *Acta Amazônica* 12(4), 677–695.

Correa, P. (1926–1969). *Dicionário das Plantas Úteis do Brasil e das Plantas Exóticas Cultivadas*, 4 Vols. IBDF, Rio de Janeiro.

Denevan, W. M. (1976). The aboriginal population of Amazonia. In *The Native Population of the Americas in 1492* (W. M. Denevan, Ed.), pp. 205–233. Univ. of Wisconsin Press, Madison.

Denevan, W. M., Treacy, J. M., Alcorn, J. B., Padoch, C., Denslow, J., and Paitan, S. F. (1984). Indigenous agroforestry in the Peruvian Amazon: Bora Indian management of swidden fallows. *Interciêcia* 9(6), 346–357.

Harlan, J. R. (1975). *Crops and Man*. American Society of Agronomy/Crop Science Society of Marica, Madison, WI.

Harlan, J. R., de Wet, J. M. J., and Price, E. G. (1973). Comparative evolution of cereals. *Evolution* 27, 311–325.

Hill, J., and Moran, E. (1983). Adaptative strategies of Wakuenai people in the oligotrophic rain forest of the Rio Negro basin. In *Adaptative Responses of Native Amazonians* (R. Hames and W. Vickers, Eds.). Academic Press, New York.

Hoehne, F. C. (1946). *Phytophysionomia do Estado do Mato-Grosso e Ligeiras Notas a Respeito da Composição e Distribuição da Sua Flora. Estudo Preliminar*. Melhoramentos, São Paulo.

Iltis, H. H. L. (1986). Maize evolution and agricultural origins. *Int. Grass Symp.*, 195–220.

Instituto Nacional de Pesquisas da Amazônia/Núcleo do Acre (1984). Levantamento de informações sobre a ocorrncia de frutas nativas no interior do estado do Acre.

Jordan, C. F. (1985). Soils of the Amazon rainforest. In *Key Environments: Amazonia* (G. T. Prance and T. E. Lovejoy, Eds.), pp. 83–94. Pergamon, New York.

Kerr, W. E., and Clement, C. R. (1980). Práticas agrícolas com conseqüências genéticas que permitiram os índios da Amazônia uma melhor adaptação às condições regionais. *Acta Amazônica* 10(2), 156–159.

Kerr, W. E., and Posey, A. D. (1984). Informações adicionais sobre a agricultura Kayapó. *Interciência* 9(6), 392–400.

Le Coint, P. (1947). *Amazônia Brasileira III. Árvores e Plantas Úteis (Indígenas e Aclimatadas)*, 2nd. ed. Ed. Nacional, São Paulo.

Lukesch, A. (1976). *Mito e Vida dos Caiapós*. Editora Pioneira de Estudos Brasileiros/Editora da Univ. de São Paulo.

May, P. H., Anderson, A., Frazão, J. M. F., and Balick, M. J. (1985a). Babassu palm in agroforestry systems in Brazil's mid-north region. *Agroforestry Syst.* 3, 275–295.

May, P. H., Anderson, A. B., Balick, M. J., and Frazão, J. M. F. (1985b). Subsistence benefits from babassu palm (*Orbignya martiniana*). *Econ. Bot.* 39(2), 113–129.

Picanço, D. B., Sousa, N. R., Clement, C. R., and Nagao, E. O. (1999). Discriminaçao de raças primitivas de pupunha (*Bactris gasipaes*) na Amazônia Brasileira com Marcadores Moleculares (RAPDs). *Genet. Mol. Biol.* 22(3), 290.

Pires, J. M., and Prance, G. T. (1985). The vegetation types of brazilian Amazon. In *Key Environments: Amazonia* (G. T. Prance and T. E. Lovejoy, Eds.), pp. 109–145. Pergamon, New York.

Roosevelt, A. R. (1991). *Moundbuilders of Amazon: Geophysical Archeology on Marajo Island, Brazil*. Academic Press, San Diego.

Salati, E. (1985). The climatology and hidrology of Amazonia. In *Key Environments: Amazonia* (G. T. Prance and T. E. Lovejoy, Eds.), pp. 18–48. Pergamon New York.

Saragoussi, M., Martel, J. H., and Ribeiro, G. A. (1988). Comparação na composição de quintais de três localidades de terra firme do estado do Amazonas. In *Ethnobiology: Implications and Applications* (D. Posey and W. L. Overal, Eds.), Proceedings of the first International Congress of Ethnobiology, 295–304. Belém, PA. Brasil.

Setz, E. Z. (1991). Animals in the Nambiquara diet: Methods of collection and processing. *J. Ethnobiol.* **11**, 1–22.

Silva-Filho, D. F., Noda, H., and Clement, C. R. (1993). Genetic variability of economic characters in 30 accessions of Cubiu (*Solanum sessiliflorum* Dunal) evaluated in central Amazonia. *Rev. Brasileira Genét.* **16**(2), 409–417.

Silva-Filho, D. F., Noda, H., Clement, C. R., and Machado, F. M. (1999). Cubiu (*Solanum sessiliflorum*): Conservação e melhoramento de etnovariedades na Amazônia. *Genet. Mol. Biol.* 22(3), 708.

Smith, N. J. H. (1980). Anthrosols and human carrying capacity in Amazonia. *Ann. Assoc. Am. Geographers* **70**, 553–566.

Smith, N. J. H. (1996). Home gardens as a springboard for agroforestry development in Amazonia. *Int. Tree Crops J.* **9**, 11–30.

Smith, N. J. H., Williams, J. T., Plucknett, D. L., and Talbot, J. (1992). *Tropical Forests and Their Crops.* Comstock/Cornell Univ. Press, Ithaca, NY.

Venturieri, G. A., and Mendonça, M. L. (1985). Cupuaçu sem semente: Histórico do aparecimento da cultivar. *Informativo Soc. Bras. Frutic.* 4(4), 12–13.

Venturieri, G. A., Rodrigues, W. A., and Menezes, J. M. T. (1986/1987). Salikisu e Kwalhakatasu, duas fruteiras dos Índios Nambiquaras com potencial para a domesticação. *Acta Amazônica* **16/17**, 1925.

CRUSTACEANS

Marjorie L. Reaka-Kudla
The University of Maryland

GLOSSARY

antennules, antennae (pl.) Anntennules are the first pair of sensory appendages on the head, and antennae are the second pair of sensory appendages on the head of crustaceans.

benthic, epibenthic Living at the bottom of the sea; *epi* = upon, thus epibenthic organisms live on top of sediment or hard substratum.

brackish Less salty than sea water.

chelate, subchelate, chela (s.), chelae (pl.) Chelate means having a pincher-like claw (usually due to an extension of the second from terminal segment beside the terminal segment, forming the claw). In subchelate forms, the terminal segment merely folds back on the second from terminal segment. The chela is the clawlike appendage.

commensal, symbiotic Living in association with (e.g., on or in) another organism.

crepuscular Active in dim light conditions, such as dusk or dawn.

dorsal, ventral Dorsal part of the organism is the "back," on the opposite side of the body from the mouth; ventral is the same side of the body as the mouth and legs for a crustacean.

extant Living.

furca (s.), furcae (pl.) Forked structure, usually associated with the telson in crustaceans.

gonochoristic Individual is either a male or a female throughout life.

hypersaline More salty than seawater.

interstitial Living among sand or silt grains; usually refers to an aquatic benthic environment, the intertidal, or in groundwater.

invertebrates Animals (usually used in reference to phyla or major taxa) that lack backbones.

macrobiota Organisms large enough to see, usually greater than 1 mm.

median, medial Along or toward the midline of the body.

phyllopod A leaflike appendage.

pleopod Abdominal appendage in crustaceans.

polyphyletic A group of species or lineages that have independent origins instead of being descended from a common ancestral group.

protogynous, protandrous Different types of hermaphroditism (in contrast to gonochoristic). Individuals in protogynous species develop into a reproductive female early in life, then change sex and become a reproductive male later in life. Individuals in protandrous species are functional males first, then females.

rami (pl.) A cuticular extension (usually relatively long and thin, can be leaflike) from the body (e.g., caudal rami = extention from the posterior segment of the body, usually paired in crustaceans).

setae (pl.) Bristle- or hairlike extensions of the cuticle that are characteristic of arthropods and related phyla.

taxon (s.), taxa (pl.) A group of organisms with shared similarities at any level of a taxonomic scheme (could be a species, genus, family, order, etc.)

telson Most posterior segment of the body in crustaceans, usually flattened and often armored.

thoracopod Appendage on the thorax (paired in crustaceans).

uniramous, biramous, triramous, polyramous An extension of the body (usually an appendage) having one, two, three, or many branches (e.g., triramous antennae have three main branches or flagellae).

vestigial An organ or structure that is present only in rudimentary form (probably lost from a previous condition).

TO UNDERSTAND THE BIODIVERSITY—and prospects for survival—of contemporary Crustacea, we need to know the number of species and higher taxa, and the range of architectural variation, habitats used, ecological roles, life history traits, and biogeographical patterns found in the group as it has developed over the last half-billion years.

I. SUMMARY AND DEFINITION OF BIODIVERSITY IN CRUSTACEA

In the last two decades of the twentieth century, it became evident to both the scientific and much of the lay public that the health of the earth's living biota—the foundation of human existence—is seriously threatened by environmental changes, mostly generated by the exponentially increasing activities of humans. This realization was made more poignant by the simultaneous recognition that our knowledge of the diversity of the world's animals, plants, fungi, and mircoorganisms is woefully inadequate even to assess the amount of diversity present and the degree of risk faced by different components of the global biota (Wilson, 1985, 1992).

The term "Biological diversity" and its shortened form "biodiversity" were introduced by Norse and McManus, (1980), Lovejoy (1980), and the National Forum on BioDiversity, held at the Smithsonian Institution (1986). Whereas "biological diversity" was first defined to include the concepts of genetic diversity (variability within species) and ecological diversity (number of species in a community of organisms; Norse and McManus 1980), the term "biodiversity" was formalized and popularized in the Proceedings of the National Forum (Wilson and Peters, 1988) to include an appreciation for the range of all of the different levels and types of biotic diversity found on earth, including the processes that generate and maintain diversity, and is thus followed here. "Diversity" sometimes implies concepts about the degree of eveness in abundance among species as well as number of species, but, for simplicity and convenience, the terms "diversity" and "biodiversity" will be used interchangeably here to indicate the number of species or higher taxa in a group and to encompass the range of structural variation, habitat characteristics, ecological roles, life history traits, and biogeographical characteristics of particular groups of organisms. The diversity of these features, including the uniqueness of certain adaptations or lineages, often can be comprehended best in the context of the ecological and evolutionary processes that generate these patterns of variation.

To properly evaluate biodiversity in Crustacea, we need to know the range of major ways of life of different groups of crustaceans, as indicated in their body plans. These architectural differences are reflected in the number of higher taxonomic groups within taxa and provide insights into the diversity of ecological processes in which the group participates.

Ecological diversity includes the range of habitats that the group occupies, whether the group lives in unique environments or species assemblages, and the diversity and importance of the ecological roles that taxa within the group play in their natural communities. For example, are the species low on the food chain (thus supporting the diversity of the rest of the community), do they affect community structure as top predators, do they determine diversity of the community by engineering the physical environment? Other important components of biodiversity include the phylogenetic history and uniqueness of lineages. Some groups of Crustacea represent the only surviving remnants of lineages that have persisted for half a billion years. Because of their early derivations, such groups are likely to be unique in their genetic, morphological, physiological, or ecological adaptations. Such a historical legacy merits special conservation attention. Similarly, the diversity of life history patterns is intimately related to genetic variability within and among populations. Some crustaceans show rapid maturation, short life spans, and massive reproductive effort, while others live for extended periods (often many years) and are morphologically and behaviorally adapted for strong interactions with other species. Some produce a few large offspring that develop directly into juvenile stages similar to the adults and do not disperse far from the parental population. Other species release huge numbers of tiny offspring, entirely different in form from the adults, that feed for extensive periods in the plankton and disperse great distances. Many of these patterns track within lineages and are correlated with body size. High variability among individuals within populations, frequently associated with wide dispersal of pelagic offspring and wide geographic ranges of species, may enable a population to persist in spite of environmental changes. However, high variability among populations (often associated with lower gene flow and isolated endemic populations) may increase the probability of losing entire genetic complements if local populations are eliminated by environmental change. Distributional patterns, particularly at the species level, are one of the most important and neglected aspects of biodiversity in most groups of organisms, including the Crustacea. The extent of the geographic range of a group, and of species within the group, are some of the most important predictors of extinction in the fossil record (Jablonski, 1991). Organisms that inhabit highly specialized habitats (especially when these habitats are geographically restricted), occupy high trophic levels, or are of large body size are likely to be vulnerable to extinction because of relatively low total population sizes and their dependence on a reliable food supply or special microhabitat (both of which fluctuate in many environments). Understanding the paleontological history of crustaceans is essential for understanding the present configuration and future prospects for crustacean biodiversity. The fossil record indicates that recoveries from major mass extinctions require about 10 million years, rendering the process of diversification irrelevant for the survival of humans and other biota over the next 500 years. However, understanding vulnerability to extinction in the past and in the present rapidly changing environment is crucial for assessing present crustacean biodiversity and sustaining it for future generations.

II. THE SETTING

A. Biodiversity in Major Global Habitats, with Special Reference to Marine Environments

Marine environments, the habitat in which crustaceans originated and which continues to sustain the greatest proportion of crustacean species, contain more higher-level biodiversity (different types of organisms represented, for example, as phyla, classes, orders, families, and genera) than other habitats because life arose in the sea, and many of these ancient lineages are still with us. Of plants, animals, and fungi large enough to see, 43 phyla live in the ocean and 28 live on land. Within the 33 animal phyla, 64% inhabit only the sea while only 5% are exclusively land dwellers, and 90% of all known classes are marine (May, 1994; Ray, 1991). It is important for humans to know and understand the diversity of these marine lineages, with all of their bizarre body plans and adaptations, in order to understand how life evolved and how we fit into it. Of more immediate urgency, marine organisms provide a critical source of protein for humans. This fact will become more important in the 21st century and beyond (especially in the world's tropical countries, which are experiencing rapid population growth and associated environmental degradation). Possibly even more significant in the long run, these marine groups provide a repository of genetic and chemical material that has been evolving, each in its own direction, for more than half a billion years. This genetic and chemical material is much more likely to be novel than the more recent, better-known lineages that invaded land.

Despite its huge expanse and the greater number of higher level groups it supports, the ocean contains fewer known species than the land. Only about 274,000 or 14% of the world's 1.9 million described species are known from the ocean (Reaka-Kudla, 1997). Of these, approximately 200,000 are macrobiota (vertebrates, invertebrates, and plants large enough to see) and 180,000 are macroinvertebrates (groups of marine invertebrates sufficiently big to see). Global freshwater environments contain approximately 44,000 total species (2.4% of all those described), 35,000 species of macrobiota and 26,000 species of macroinvertebrates. There are about 950,000 insects, 80,000 noninsect and noncrustacean (mostly chelicerate) arthropods, and 57,000 other terrestrial macroinvertebrates. Globally, about 100,000 species live symbiotically on or in other organisms (5.3 % of all described species); roughly 73,000 of these probably are macrobiota (mostly invertebrates). If we exclude the noncrustacean terrestrial arthropods (chelicerates, insects, and their relatives) because of their unique specializations for life on land, there are approximately 336,000 described macroinvertebrate species in the marine, freshwater, terrestrial, and symbiotic environments. Of these, marine macroinvertebrates constitute 53.6%, freshwater macroinvertebrates 7.7%, terrestrial macroinvertebrates 17.0%, and symbiotic macroinvertebrates

21.7% of the total species. We can use these figures to estimate the numbers of crustacean species that occupy different types of habitats.

There are about 43,000 species of described crustaceans—2.3% of the 1.9 million described species and 12.8% of all macroinvertebrates. If crustaceans inhabit major types of environments in about the same proportioins as other nonchelicerate and noninsect macroinvertebrate groups, probably about 23,000 described species of crustaceans live in marine and brackish coastal waters (figures rounded to nearest thousand); this represents about 12% of all species of marine macrobiota. Similarly, if crustaceans have invaded fresh water to about the same degree as other nonchelicerate and noninsect invertebrates, at least 3400 crustacean species likely inhabit fresh waters (rounded to nearest hundred); this represents about 10% of all described freshwater macrobiota. If crustaceans are about equally adapted to land as are other nonchelicerate and noninsect invertebrates, about 7300 crustaceans are terrestrial. This is less than 1% of the global terrestrial macrobiota. Since about 22% of all macroinvertebrate species are symbiotic, probably about 9,500 crustacean species are estimated to be symbionts (about 13% of global symbiotic macroinvertebrates). In fact, parasitism is rampant in some groups of crustaceans. Certainly these figures are underestimates—most likely great underestimates—of the true numbers of species in aquatic and particularly marine environments because these environments are so poorly known. If only 10 to 20% of all crustacean species have been described (discussed later), then there likely are about 200,000 to 450,000 total (known and unknown) species of crustaceans on earth.

The following observations suggest that the numbers of crustacean species, as well as those of other marine phyla, usually are underrepresented. Based on samples of 1597 species of soft-bottom marine macrofauna off the east coast of the North America (255 to 3494 m), Grassle and Maciolek (1992) calculated that the global deep-sea fauna, primarily because of the huge area it occupies, may include as many as 10 million species (mostly polychaete annelids, crustaceans, and mollusks). Most of these species are small and rare (90% of the species sampled comprised less than 1% of the individuals; 28% of the species in the entire fauna were collected only once). Reaka-Kudla (1997) calculated that global coral reefs contain between 1 and 5 million total species (though only about 100,000 described species).

Experts have estimated that about 21% of global crustacean species have been described and that about 26% of global molluscan species are known (Systematics Agenda 2000, 1994). These phyla represent the two most commercially important and thoroughly documented invertebrate groups. Only 10% of the relatively well-known isopod crustaceans collected from shallow sediments in southern Australia had been described (Poore and Wilson, 1993). Because of the great variation in the extent to which different regions of the oceans have been sampled, these scientists estimated that only 5% of marine invertebrates are known from the oceans overall. Calculations suggest that about 10% of the species on global coral reefs have been described. Scientific descriptions exist for about 17% of the species of algae. The microbiota is even more poorly known. Only 1%, 1 to 10%, 4 to 7%, and 2 to 3% of the total species of the viruses, bacteria, fungi, and nematodes have been described.

Although the ancient major groups (phyla, classes) of marine organisms are more different from each other than those on land, species within marine invertebrate groups often look alike, leading to underestimates of their true numbers. Recent molecular and genetic studies have shown that described species frequently include several different, previously unrecognized species (Knowlton, 1993).

Major new groups of organisms are still being discovered in the sea, whereas this is rare on land. At least three new invertebrate phyla (the largest category of related organisms) and many major new groups (classes, orders, families, genera, which represent increasingly smaller groupings of related organisms) of deep-sea organisms, picoplankton, marine viruses, and archaebacteria (a new kingdom of one-celled organisms that often live in extreme environments) have been described over the past 20 years. In the crustaceans, a new class, the primitive Remipedia, were discovered in caves in the Bahamas and described only in 1981. A new subclass of the Maxillopoda, the Tantulocarida (tiny ectoparasites of deep sea copepods, ostracods, and isopods) was erected in 1983. In the relatively well known malacostracan crustaceans, the Order Spelaeogriphacea was not discovered until 1957, the Order Amphionidacea recognized only in 1973, and the Order Mictacea established as late as 1985 (Schram, 1986). Crustacean biology, and marine biology in general, remains a realm of discovery.

Aquatic environments are difficult to see, sample, and study. The number of scientists who study invertebrates and the amount of scientific effort devoted to invertebrates other than insects is two orders of magnitude less than that devoted to vertebrates and one order of magnitude less than that devoted to plants (Gaston and May, 1992). In addition, tropical environments, far from the location of most biologists, have been much less studied than those at higher latitudes; 80% of ecological researchers are based in North America and Europe in comparison to 7% in Latin America and tropical Africa. Identification of species in aquatic habitats is further reduced because of the inability to observe them easily and the frequent requirement for Scuba diving and extensive logistic support.

Moreover, most crustacean and other aquatic taxa are small in body size. Many aquatic species live within crevices, holes, or burrows in coarse (rocks, coral, rubble) and level (mud, sand) substrata. Many are crepuscular or nocturnal to minimize fish predation (Dominguez and Reaka, 1988), making them difficult to see, collect, and study. Small organisms of any environment always are poorly known and studied, but this is more extreme in aquatic environments, particularly in tropical marine environments, because of the remote localities and the difficult conditions for observing and collecting.

It seems likely that major advances in our knowledge of the extent of crustacean biodiversity will continue into the next centuries. These advances are imperiled, however, by a crisis in the number of systematists and taxonomists being trained (Feldmann and Manning, 1992; Reaka-Kudla et al., 1997; Wilson, 1985). Although we know a fair amount about the morphology, ecology, behavior, life history, and development of many of the crustacean groups, we know astonishing little about the facets of their biology that are most critical for assessing their biodiversity status, particularly the distribution and abundance patterns of individual species and the life history features that render the species vulnerable or resistant to extinction. More distressing, we are incapable of determining these aspects of the biology of given taxa without a reliable foundation of systematics and taxonomy. The understanding and conservation of crustacean biodiversity will depend on the implementation of major new funding and training programs that are directed toward systematics and taxonomy, and toward natural history museums which hold the collections necessary for such comparative studies.

B. Ecology of the Arthropod Radiation

Between 3 and 4 billion years ago, biopolymers (proteins, polysaccharides, nucleic acids), which had been formed as a result of physical processes, became enclosed in membranes, and the resulting cells developed the ability to replicate themselves. These duplicated cells became organized into multicellular organisms that could move about on their own about a billion years ago. An accelerating explosion of biotic diversity occurred sometime in the Precambrian (perhaps 700–800 million years ago), but fossils remnants of this early proliferation are few, having been either poorly preserved or destroyed in subduction zones beneath migrating continents. By the beginning of the Paleozoic era (about 570 million years ago [mya]), almost all of the major phyla of animals already were present (by then being larger and more amenable to fossilization or preserved in continental rocks that are still available). The Cambrian marine environment (approximately 570–500 mya) was studded with sessile epibenthic forms such as stromatolites (aggregations of unicellular organisms, their secretions, and sand or other sedimentary particles); sponges and other now extinct spongelike forms; upright branched and laminar bryozoans ("moss animals"); filamentous now extinct organisms called graptolites, brachiopods ("lamp shells"), and other sessile organisms with ciliated feeding arms; various types of cnidarians (hydroids, anemones, primitive types of corals, jellyfish); stalked crinoids (sea lillies) and other primitive echinoderms; and the sessile life stages of primitive forms of chordates (our own ancient forebearers). Most of these sessile organisms fed on suspended particles near the seafloor.

Various other organisms crawled on the seafloor, many of them having developed segmentation (and other architectural features that allowed them to move more rapidly and feed more effectively) as well as external armor that protected them from the increasing tendency for their ilk to eat one another. In addition to many lineages (often with bizarre features unlike any organisms known today) that perished in the race for motility and are known only as fossils, the motile benthic fauna included the early mollusks (whose followers would diversify during the Paleozoic into major grazers, suspension feeders, burrowing sediment feeders, and large extraordinarily active predators). Also present were diverse groups of early echinoderms (that would become grazers, burrowing sediment feeders, and slow-moving predators as the Paleozoic progressed) and the early filter-feeding chordates (many of which would develop into large, increasingly active and important predators as jawless and then jawed fish during the early to mid-Paleozoic). At least as important for biodiversity in the long run, the Precambrian benthic marine environment abounded with a diverse array of unsegmented and segmented worms that provided the prototype for the development of the joint-legged animals, the arthropods.

Whether the arthropod phyla (uniramians [centipedes, millipedes, insects], onycophorans, tardigrades, chelicerates [horseshoe crabs, scorpions, spiders, ticks], pycnogonids, trilobites, pentastomids, and crustaceans) arose from only one group or independently from several groups of the early segmented annelid worms is hotly debated. The annelids and arthropod groups originated in the Precambrian, and all of the major phyla were distinct by the time they were fossilized. The crustaceans, trilobites, and chelicerates appeared first in the fossil record, followed at a later time by the uniramians. Some experts have concluded, on the basis of functional morphology and developmental patterns, that the major arthropod groups probably all arose independently, and that the Crustacea are not closely related to the uniramians, trilobites, or chelicerates. Other scientists have argued, on the basis of morphology, that the trilobites are ancestral to both the chelicerates and crustaceans. New evidence from RNA and DNA sequences has brought forth various views, including the idea that the crustaceans are more closely related to the uniramians than to other arthropod groups. The question of how the arthropod groups evolved, then, remains far from solved at present.

Many workers concur that the onycophorans (velvet worms) and the uniramians have a common origin. Some workers have suggested a possible connection between two Cambrian marine fossils and the onycophorans, who are represented today by 80 species in warm moist terrestrial environments. The first myriapods (centipedes, millipedes) were fossilized in the Silurian (440–405 mya) and probably they or their ancestors were marine. Terrestrial myriapods and wingless insects are known from the Devonian (405–360 mya), and winged insects had arisen by the Carboniferous (360–290 mya). These underwent a continuing spectacular radiation that resulted in almost a million species of described insects that inhabit almost all terrestrial and many aquatic environments today. In addition to 32 living orders of insects, at least 10 other orders arose and went extinct in the late Paleozoic and early Meszoic.

About 400 species of tardigrades (water bears) are known from marine (intertidal to deep-sea) environments, freshwater, hot springs, and semiterrestrial environments in plants and humus; the most primitive group is marine. A single fossil is known from Cretaceous (135–65 mya) amber. These minute organisms (most 0.1–0.5 mm, a few to 1.7 mm) occupy benthic, interstitial, water film, commensal, and parasitic habitats. They also have been been collected from some of the driest habitats on earth, the Antarctic dry valleys; they are capable of being freeze-dried, after which they resume metabolism when rehydrated.

Represented by the xiphosurans (horseshoe crabs) and the eurypterids (sea scorpions), the early chelicerates were marine and had appeared by the Cambrian. Eurypterids and xiphosurans were crawling scavengers and predators, crushing food between the bases of their legs. The anterior appendages (chelicerae) of some eurypterids were well developed and clawed. Many eurypterids were small but some grew to very large sizes (up to 3 m). These marine forms flourished during the Ordovician, Silurian, and Devonian (500–360 mya) and were abundant until the Permian (290–245 mya), after which the eurypterids became extinct and the xiphosurans were greatly diminished (only five species of horseshoe crabs remain today). Some eurypterids apparently invaded fresh water and may have become semiterrestrial. Related lineages, including the scorpions (the most primitive living arachnids) were present in the Silurian and invaded land very early, at least by the Carboniferous. Diversification of the Chelicerata. resulted in a total of about 80,000 described, mostly terrestrial species living today.

The pycnogonids (sea spiders), considered a sister group to the Chelicerata, are known from the Devonian onward. They are represented by about 1000 species and 86 genera today, all marine. They are suctorial predators and occupy all oceans down to about 7000 m. Most are small (<10 mm), though the bodies of some deep sea forms are 20 to 30 mm with leg spans of up to 700 mm. The "protonyphon" larva of some pycnogonids and the "trilobite" larva of some xiphosurans share some developmental features with the "protaspid" larva of trilobites.

The trilobites, along with the chelicerates, were major components of the motile benthic fauna of the early Paleozoic. The trilobites were exceedingly abundant during the Cambrian and Ordovician (600–440 mya) and persisted in somewhat lower numbers until the end of the Permian (245 mya), after which they became extinct. All of the 4000 species were marine. They were major scavengers and predators, crushing food between the bases of their legs, which operated as gnathobase jaws. Some were possibly sediment or suspension feeders. Most species were small (10–30 mm), some were moderately sized (30–60 mm), and a few giants (600–700 mm) are known. Most species were benthic and epifaunal or shallow infaunal, but some small (<10 mm) forms were planktonic.

The pentastomids, wormlike parasites that inhabit the respiratory system of terrestrial vertebrates, include about 95 species and are thought to be related to crustaceans. They may have been a relatively late offshoot of early arthropod lineages, developing with or after the radiations of terrestrial vertebrates in the Carboniferous (amphibians), Permian (reptiles), and Mesozoic (reptiles, birds, mammals).

In the midst of this evolutionary theater rose the crustaceans. We shall return to the historical development of crustacean diversity after we are introduced to the major characters.

III. BIODIVERSITY OF FOSSIL AND MODERN CRUSTACEA

A. The Major Groups of Crustacea

1. Centers of Taxonomic Diversity

At least five major radiations in ancestral crustacean lineages produced the remipedes, cephalocarids, branchiopods, maxillopods, and malacostracans (Table I). Scientists believe, either because of their primitive body forms or their fossil record, that all five major groups have ancient origins extending back at least as far as the beginning of the Paleozoic. The remipedes, cephalocarids, and branchiopods probably are impoverished relicts of ancient great lineages that did not fossilize well. Today, however, these three groups (along with a number of primitive malacostracan lineages) are marginalized in hypersaline, brackish, and fresh water (sometimes hot springs or ephemeral pools), caves, groundwater, or other isolated, seemingly suboptimal habitats. Collectively, these three classes contribute only 14% of the diversity in orders, 5% of the diversity in families, 2% of the generic diversity, and 2% of the species diversity within modern crustaceans. The remaining 2 classes—the maxillopods and especially the malacostracans—are the dominant groups in most major marine environments today and are responsible for much of the diversity in major body plans (e.g., as represented in subclasses; Table I). The malacostracans alone constitute over half (53%, 59%, and 55%) of all present and past (fossilized) crustacean families, genera and species, respectively. The major groups of malacostracans have been subject to considerable turnover in the fossil record, however, since the lineage is ancient and many of their major evolutionary experiments (body plans at the level of order) have been lost to extinction. Whereas 80 to 100% of all the orders of branchiopods and maxillopods ever known are still extant, only 59% of malacostracan orders are still living. If their order did survive background extinction and the great mass extinctions (e.g., at the ends of the Paleozoic and the Mesozoic), however, malacostracan families have persisted (or are currently diversifying) at a higher rate (88% of all known families are still extant) than those of the maxillopods (76% living) and branchiopods (74% extant). Though only 87% of branchiopod genera still exist, very high percentages of genera in both the malacostracans (97%) and maxillopods (96%) are alive today.

Within the maxillopods, centers of taxonomic diversity are found in the copepods, ostracods (seed shrimp), and, to a lesser extent, cirripeds (barnacles). Copepods include 38% of the orders, 50% of the families, 54% of the genera, and 49% of the species in the maxillopods. Ostracods constitute 29% and 38% of the orders and families, and 36% and 44% of the genera and species in the maxillopods. Barnacles comprise 19% and 11% of the orders and families, and 10% and 6% of the genera and species in the group. Both ostracods and barnacles are ancient lineages whose fossil records extend back at least to the dawn of the Paleozoic.

Within the malacostracans, peracarids and eucarids dominate the group in terms of taxonomic diversity. Peracarids include 37% of the orders and 62% of the families, and at least 56% of the genera and 53% of the species, in the Malacostraca. The eucarids make up only 11% of the orders and 28% of the families, but at least 37% of the genera and 44% of the species within the group.

2. Architectural Diversity

To understand the diversity of Crustacea, it is necessary to come to grips with their major ways of life as reflected in their body plans. The range of morphological diversity among crustaceans is remarkable, considerably greater than that of the insects. Primitively, each segment supported a pair of appendages. Crustacean appendages have multiple joints along their length and are considered to have evolved from a biramous condition (divided into two branches), although both head and trunk appendages have lost one side of the branch in some groups. The primary mode of evolutionary change in the early diversification of crustaceans was by molding segments into functional units along the length of their body and reducing various morphological features, including both segments and appendages. The resulting fundamental body regions included a five segmented head (cephalon) and a long, segmented trunk that is usually divided into a thorax and abdomen.

The head was formed by telescoping five segments and their associated appendages into a fused unit that bears two pairs of antennae (a distinguishing character-

TABLE I

Minimum Numbers of Taxa for the Five Major Crustacean Evolutionary Radiations

Class	Subclasses	Orders	Families	Genera	Species
Class Remipedia	—	2 (1 living), 4%	1 (1 living), 0.1%	6 (5 living), 0.1%	11 (10 living), <0.1%
Class Cephalocarida	—	2 (1 living), 4%	2–3 (1–2 living), 0.2%	At least: 5 (4 living), 0.1%	At least: 11 (10 living), <0.1%
Class Branchiopoda	—	4–5 (4 living), 7%	At least: 35 (26 living), 4%	At least: 108 (94 living), 2%	At least: 836 (821 living), 2%
Class Maxillopoda	6 living	21 (18 living), 38%	355 (270 living), 43%	At least: 2162 (2077 living), 39%	At least: 18,388 (18,293 living), 43%
	Ostracoda	6 (3 living), 29%	136 (55 living), 38%	At least: 776 (695 living), 36%	At least: 8081 (about 8000 living), 44%
	Mystacocarida	1 living, 5%	1 living, 5%	2 living, 0.1%	10 living, 0.1%
	Copepoda	8 living, 38%	177 living, 50%	1167 living, 54%	About 9000 living, 49%
	Branchiura	1 living, 5%	1 living, 0.3%	4 living, 0.2%	150 living, 1%
	Tantulocarida	1 living, 5%	1 living, 0.3%	2 living, 0.1%	5 living, <0.1%
	Cirripedia	4 living, 19%	39 (35 living), 11%	At least: 211 (207 living), 10%	At least: 1142 (1138 living), 6%
Class Malacostraca	2–7 (usually 5) living	27 (16 living) (3 extinct groups unassigned to subclasses), 48%	431 (383 living), 53%	At least 3175 (3093 living), 59%	At least: 23,890 (23,800 living), 55%
	Phyllocarida	5 (1 living), 19%	13 (2 living), 3%	At least: 55 (5 living), 2%	At least: 65 (13 living), 0.3%
	Hoplocarida	3 (1 living), 11%	At least: 19 (13 living), 4%	At least: 78 (68 living), 2%	At least: 413 (400 living), 2%
	Syncarida	3 (2 living), 11%	10 (6 living), 2%	At least: 37 (33 living), 1%	At least: 119 (115 living), 0.5%
	Peracarida	10 (9 living), 37%	268 (255 living), 62%	At least: 1793 (1775 living), 56%	At least: 12,726 (12,706 living), 53%
	Eucarida	3 living, 11%	121 (107 living), 28%	At least: 1213 (1212 living) (fossils mostly not quantified), 38%	At least: 10,567 (10,566 living) (fossils mostly not quantified), 44%
Totals	—	56–57 (40 living)	At least: 824 (681 living)	At least: 5457 (5273 living)	At least: 43,136 (42,934 living)

In cases where number of taxa is controversial, the lower number was used in the total. In many cases, the number of certain taxa (especially genera and species) were not summarized in sources, so minimum numbers were estimated from the number of higher taxa (e.g., in a group with five families for which data on number of genera were not summarized, it was assumed that at least five genera [one per family] and at least five species [one per genus] were present). Therefore, these data probably considerably underestimate the number of genera and particularly the number of species in these classes (although total species exceed those previously estimated by other authors, e.g., Brusca and Brusca; 1990, who did not include paleontological data). Percentages horizontally allied with classes (bolded) represent the total number in that taxon compared to the number for all Crustacea. Percentages horizontally aligned with subclasses (not bolded) represent the total number in that taxon compared to the total number in that class. Percentages are based on total (fossil and living) taxa and are rounded to the nearest whole number (except when <1%).

istic of crustaceans), the mandibles (jaws) that flank the mouth, and two pairs of accessory food handling appendages, the maxillules and maxillae. The antennules (first pair of antennae) are biramous in the primitive remipedes. However, the antennules are uniramous in the cephalocarids, all of the branchiopods, all of the maxillopods (where present), and some of the eumalacostracans, while those of the primitive phyllocarid malacostracans and some of the eumalacostracans are biramous and those of stomatopods triramous, raising the question of whether or not the antennules of crustacean ancestral groups all were biramous. The antennules are reduced or lost in the parasitic branchiurans and tantulocaridans and in many sessile adult cirripeds. The second antennular appendages (antennae) are biramous in most crustaceans but are uniramous in anostracans, some copepods, leptostracans, some but not other fossil phyllocarids, and some eumalacostracans. The antennae are reduced in the parasitic branchiura (modified for attachment) and lost in the notostracans, tantulocarids, and sessile adult cirripeds.

One or more of the anterior thoracic segments may be fused to the head (remipedes, isopods, amphipods), or the anterior thoracic appendages (maxillipeds) may be functionally associated with the mouthparts of the head (fossil kazacharthran branchiopods; fossil lipostracans; ostracod, mystacocaridan, and copepod maxillopods; stomatopod and possibly palaeostomatopod hoplocaridans; anaspidacean syncarids; mysidan, lophogastridan, tanaidacean, spelaeogriphacean, cumacean, mictacean, isopod, and amphipod peracarids; stomatopods and many eumalacostracans).

Reflecting very early divergence among groups of crustaceans, the division of the trunk into units that functioned in feeding (the anterior appendages), reproduction, and locomotion (often with accessories that aided respiration) was much more variable among major crustaceans groups than the head region. The wormlike remipedes usually are considered to represent the most primitive crustacean morphology. Their oarlike appendages are undifferentiated throughout the long trunk, followed by an anal segment with simple caudal rami. The trunk of the cephalocarids, however, is divided into 8 thoracic segments, 11 abdominal segments, and an anal segment with caudal rami. The trunk region of both the branchiopods and maxillopods is variable among groups. Several groups with modified morphology possess simply a reduced trunk (cladocerans, conchostracans, ostracods, mystacocarids, and adult tantulocarids). A thorax is present in the anostracans and notostracans (usually 11 segments), extinct kazacharthrans, copepods (6 segments), branchiurans (4 segments), tantulocaridans (6 segments in juveniles), cirripeds (6 segments, reduced), and malacostracans (8 segments). The number of segments included in the abdomen is variable within several groups of branchiopods (anostracans, notostracans, fossil kazacharthrans, conchostracans) and in juvenile tantulocarids and ascothoracidan barnacles. The abdomen is reduced or lost, or segmentation obscured, in groups that are modified for specialized ways of life (cladocerans, the parasitic branchiurans and tantulocaridans, and most cirripeds, especially the parasitic forms). The primitive phyllocarid malacostracans all possess an abdomen with seven segments, in addition to a terminal anal segment with caudal rami. Adult hoplocaridans and eumalacostracans have lost one of the abdominal segments, however, and bear six abdominal segments plus an additional flattened anal segment termed a telson. Morphological and developmental evidence of a vestigial seventh abdominal segment (either at the front or the posterior of the abdomen) in the stomatopod hoplocaridans and in the mysidan, tanaidacean, and possibly the isopod eumalacostracans reflects another evolutionary tendency toward reduction of segments.

Trunk appendages in the remipedes are oarlike but different from the leaflike or phyllopodous appendages of the primitive branchiopods. Leaflike or phyllopodous appendages predominate in the cephalocarids, all of the branchiopods, and in the primitive malacostracans, including a number of extinct groups and the modern leptostracans. However, greatly modified or reduced numbers of trunk appendages (especially on posterior trunk segments) characterize the maxillopods (e.g., the ostracods, mystacocarids, and copepods). Parasitic forms (e.g., the adult rhizocephalan barnacles) often lose all appendages. The copepods, branchiurans, and juvenile tantulocarids all have biramous thoracic legs. In the malacostracans, however, some of the thoracic appendages are uniramous and some biramous in stomatopods, and all eight pairs of the thoracopods usually are uniramous (with some of these enlarged and clawed) in the eumalacostracans. Even some of these groups, however, show reduction in development of appendages toward the posterior of the thorax. Branched or layered structures arising from the base of the thoracic legs (epipods) serve as gills in the hoplocaridans and eumalacostracans.

Posterior trunk or abdominal appendages are reduced in most cephalocarids, branchiopods, and maxillopods (notostracans, conchostracans) or absent (cephalocarids, anostracans, cladocerans, ostracods, mystacocarids, copepods, branchiurans, tantulocarids, and cirripeds). Like the remipedes, however, most of the malacostracans bear pleopods on the abdomen that often function in swimming (and sometimes facilitate mating). These may be present or absent in different fossil phyllocarids and are present, but reduced toward the posterior, in the leptostracans. Hoplocarids have five pairs of well-developed pleopods that support tufty gills (palaeostomatopods, stomatopods). Some peracarid eumalacostracans (isopods) also bear breathing structures (gills or tracheae) on their pleopods. The eumalacostracans have pleopods, but their number and size often is reduced (particularly toward the posterior) or may be absent (anaspidaceans, some mysidans, some thermosbaenaceans, some cumaceans, and some anomurans). The sixth abdominal segment is fused to the telson in modern (but not ancient) tanaidaceans, as well as some cumacean and isopod eumalacostracans. The hoplocaridans and most groups of eumalacostracans have a well-developed "tailfan" that is composed of uropods (bladelike appendages on the sixth abdominal segment) and a telson (a flattened and often armored posterior segment that may have one or more terminal projections but never bears caudal rami).

The position of the gonadal duct openings (gonopores) on either the ventral part of the body or the bases of the legs varies consistently among groups of crustaceans, reflecting the ancient divergence of different taxa. The female gonopore opens on the sixth thoracic segment, and the male gonopore on the eighth thoracic segment, in all malacostracans.

Both simple (ocelli) and compound (faceted) eyes are present in most taxa. One medial ocellus frequently occurs in larvae, and one or more ocelli are found in adults of some groups of branchiopods and maxillopods such as the anostracans, notostracans, cladocerans, ostracods, mystacocarids, copepods, and branchiurans. Only ocelli (or ocellus-like structures) are present in the adults of the extinct kazacharthrans, some cladocerans, mystacocarids, and copepods. Compound eyes often are positioned on stalks, as in the anostracan branchiopods, most malacostracans, including many phyllocarids (extinct canadaspidans, leptostracans), the hoplocarids (extinct aeschronectids, stomatopods), extinct belotelsonideans, and most eumalacostracans (extinct palaeocaridaceans, some anaspidaceans, mysidans, lophogastridans, thermosbaenaceans [nonfunctional], some tanaidaceans [on lobes], spelaeogriphaceans [on lobes], cumaceans [on a medial lobe], mictaceans [nonfunctional on lobes], some isopods [on lobes], euphausiaceans, amphionidaceans, dendrobranchiate shrimps, procaridid shrimps [reduced], caridean shrimps, stenopodidean shrimps, thalassinidean shrimps, astacideans, palinurans, anomurans, and brachyurans). Eyes are reduced or absent in members of primitive lineages, sessile lineages, or species that inhabit caves or the deep sea (the remipedes, cephalocarids, many ostracods, tantulocarids, adult cirripeds, tanaidaceans, some spelaeogriphaceans, many cumaceans [especially females], some mictaceans, some isopods, some amphipods, amphionidaceans, procaridid shrimps, some caridean shrimps, some stenopodid shrimps, and some astacideans). The branchiuran maxillopodans have a particularly intriguing pair of compound eyes that project internally into a blood sinus suspended on a stalk composed of the optic nerves. Muscles move the eyes within the blood sinus.

Either a cephalic shield or a carapace may cover the head and part or all of the thorax. Remipedes, cephalocarids, copepods, and tantulocarids bear a cephalic shield, while a "folded" or "bivalved" carapace characterizes the notostracans, extinct kazacharthrans, cladocerans, conchostracans, ostracods, branchiurans, cirripeds, phyllocarids, hoplocaridans, and most eumalacostracans. However, these structures are reduced or absent in other groups (anostracans, mystacocaridans, syncarid and most peracarid eumalacostracans). The carapace of cirripeds is modified into a fleshy, saclike mantle that may envelope most of the body and, in thoracicans, is covered with calcareous plates; eggs are brooded in the mantle cavity for various lengths of time. Conchostracans, most cladocerans, and some ostracods use their carapace as a brood chamber; one cladoceran broods eggs in its molted carapace. Thermosbaenacean females possess a very enlarged, thin, blood-filled carapace that functions both as a brood chamber and for gaseous exchange. The enlarged, very thin carapace in female amphionidaceans may serve a similar function, though this has not yet been examined. The carapace extends laterally in decapods to enclose the base of the legs and gills, increasing the efficiency of water flow over the gills.

B. The Rise of Crustacean Diversity in the Fossil Record

All of the early continents, including Gondwana (modern South America, Africa, Madagascar, India, Australia, Antarctica), Baltica (Scandinavia and northwestern Europe), Laurussia (North America), Siberia, Kazakhstania, and China, girdled the earth in tropical and subtropical locations. No land extended above 60° N or below 60° S during the Cambrian (570–500 mya), allowing mild warm conditions in most coastal areas. Developing from segmented wormlike forms, the earliest crustaceans must have developed sets of pronglike or leaflike legs to shuffle along the surface of the soft marine sediments. In addition, they secreted a tough, often calcified cuticle to protect them from the other benthic predators (nautiloid molluscs and trilobites, which were the most abundant benthic predators and scavengers in most marine habitats during the Cambrian). The earliest fossilized crustaceans included three orders of phyllocarid malacostracans (phyllopodous detritus feeders, scavengers, and generalized predators known from the Lower Cambrian); the thoracican goose barnacles (suspension-feeding maxillopods known from the Middle Cambrian); and the ostracods (scavenging, detritus-feeding maxillopods known from the Cambrian onward). Following the recent disovery of the predaceous Remipedia, which exhibit a very primitive arrangment of segmentation and appendages, many workers concluded that the remipedes, though not readily fossilized, represent one of the probable forms of the earliest crustaceans and likely were present in the Cambrian. In addition, the primitive morphology of the ascothoracidan barnacles suggests that they also may have been present (though unfossilized) in the very early Paleozoic. Of these early groups, the canadaspid phyllocarid malacostracans did not persist into the Ordovoician.

During the Ordovician (500–440 mya), part of the huge island continent of Gondwana slipped beneath what was then the south pole, but the climate remained generally warm and mild, with the northern reaches of Gondwana and most of the island continents situated in tropical or subtropical waters. Xiphosuran (horseshoe crabs) and eurypterid (sea scorpions) chelicerates joined the exceedingly abundant, ravenous trilobites on the seafloor. Nautiloids continued to diversify morphologically; jawless fish (encased in protective outer layers of bone and still not very motile) proliferated. The ostracod crustaceans underwent their greatest radiation, and phyllocarid malacostracans (hymenostracans, archaeostracans) were abundant; the hymenostracans, however, disappeared after the Ordovician. A mass extinction closed the Ordovician.

The Silurian (440–405 mya) also was characterized by mild climates (except in the parts of Gondwana that rested over the pole), and a circumtropical current wound its way among the island continents and reefs composed of stromatoporoid sponges, rugose corals, and other sessile fauna. Eurypterids flourished, trilobites persisted in lesser numbers, myriapods and primitive arachnids (probably marine) arose, and jawed fish made their appearance. Thoracican goose barnacles (attached to a eurypterid) are recorded, ostracods flourished, and archaeostracan phyllocarid malacostracans were abundant.

During the Devonian (405–360 mya), predation from the flourishing eurypterids, increasingly motile jawed fishes, and nautiloid cephalopod mollusks dominated benthic ecology. Trilobites still occupied many benthic marine environments. Pycnogonids were present. In an era of mild climates and inland seas, fish and invertebrate groups radiated into freshwater environments. A circumtropical current coursed between Gondwana and the island continents. It was in these warm, sheltered, shallow coastal waters laden with archipelagos that crustaceans began their great radiations. The fossil record shows that the archaeostracans (phyllocarid malacostracans) reached their greatest diversity, and the predaceous hoplocarids (palaeostomatopods, archaeostomatopods) made their first appearance. The ostracods began their second great radiation. The boring acrothoracican barnacles and the first predaceous, lobster-like decapod malacostracans appeared (*Palaepalaemon*, allied either to the astacid crayfish and lobsters or palinurid lobsters). Detritus-feeding and scavenging conchostracan branchiopods originated, invading fresh water, and the detritus-feeding lipostracans (allied to the cephalocarids or branchiopods) flourished in hot springs. Fossils and primitive morphology suggest that other primitive detritus-feeding branchiopods (notostracans, anostracans) likely were present too. Another mass extinction impacted faunas during the Late Devonian.

The taxonomic, morphological, and ecological diversification of the Crustacea accelerated during the Carboniferous (Mississippian, Pennsylvanian; 360–290 mya) of the Late Paleozoic as the marine chelicerates and trilobites began their decline. The continents were beginning to congeal, setting the stage for the global climatic changes that likely contributed to the greatest mass extinction of all time at the end of the Paleozoic. Gondwana had largely completed its migration beneath the south pole and emerged to begin to fuse with Laurussia and Baltica (modern North America, Scandinavia, and northwestern Europe) over the equator. Though the southern reaches of Gondwana (modern South America) were still plagued with bouts of glaciation, the warm moist equatorial continents were awash with inland seas and swamps. Land plants had multiplied in diversity and biomass, their seed ferns, scale trees, and conifer trees producing great forests that decayed in coal swamps or washed out onto the continental shelves, enriching coastal sediments. Insects took flight (including some huge species), amphibians underwent their greatest radiation, and the first reptiles appeared. Marine organisms proliferated in the enriched coastal sediments. Sharks and jawed boney fishes were abundant and active. Nautiloids declined, replaced by the emerging ammonoid cephalopod mollusks. Remipede crustaceans are recorded from the Lower Pennsylvanian. Conchostracan branchiopods radiated anew. Notostracan branchiopods probably were present. Acrothoracican barnacles are known from their borings in clam shells, and lepadomorph thoracican barnacles are recorded from this period. Early malacostracan lineages (the scavenging archaeostracan and hoplostracan phyllocarids; the scavenging aeschronectids and predaceous hoplocarids; the scavenging belotelsonidans, waterstonellideans, eocaridaceans, and paleocaridaceans) were exceedingly abundant and dominated the shallow marine fauna. Pygocephalomorphan and lophogastridan mysidaceans came on the scene, as well as most of the peracarid malacostracans (tanaidaceans, spelaeogriphaceans, cumaceans, isopods, and probably other unfossilized groups). Definitive palinuran decapods appeared, and the earliest lineages leading to the brachyuran crabs evolved. Several of the early malacostracan groups (hoplostracan phyllocarids, aeschronectid and palaeostomatopod hoplocarids, belotelsonids) disappeared after the Carboniferous as the peracarid and eucarid malacostracans took center stage.

The Permian period (290–245 mya) brought one of the most significant geomorphological events in the history of the earth. All of the previous continental landmasses fused into a single gigantic continent, Pangaea, between 260 and 250 mya. The collision of Gondwana, Laurussia, Baltica, Kazakhstania, and Siberia is still visible in the mountain belts stretching across northern South America, northwestern Africa, eastern North America (the Appalachians), Britain and Scandinavia, and Eurasia (the Urals and other ranges). By 250 mya, the ancient island continent of China also had joined the supercontinent. Pangaea now straddled the equator, anchored by Siberia near the north pole and Antarctica over the south pole. The circumtropical seaway was closed. The elevated landmasses produced a more extreme, cooler, continental climate than had been present throughout most of the Paleozoic (whether or not this was solely responsible for the mass extinctions is debated). In the cool, dry conditions, reptiles radiated, replacing the previously dominant amphibians, and the first mammal-like reptiles appeared. Many organisms, including worms, heart urchins, sea cucumbers, and crustaceans, mined the organic riches in the coastal sediments, throwing up sediment that buried and clogged the feeding structures of the epibenthic sessile filter feeders. Conchostracan branchiopods, ostracod maxillopods, primitive malacostracans (archaeostracan phyllocarids, the first leptostracan phyllocarids, waterstonellideans, eocaridaceans, palaeocaridacean syncarids, and pygocephalomorph mysidaceans), at least several peracarid malacostracans (tanaids, cumaceans, isopods), and the first definitive astacid (crayfish- or lobster-like) eucarid malacostracans thrived in benthic marine environments. Brachyurans likely were present. The Paleozoic era ended, however, with the loss of 96% of all species. Many crustacean groups did persist through the Permo-Triassic biotic catastrophe, though, and underwent reradiation during the Mesozoic.

The early Triassic was a relatively barren environment both on the land and in the sea. The landmasses were high, cool, and dry, with widespread deserts. In coastal waters, no corals or reef communities are known. All of the trilobites, eurypterids, most of the nautiloid cephalopods, and many of the primitive fishes had been lost. All of the primitive phyllocarid malacostracans, except for the newly evolved leptostracans, were gone. All of the early hoplocarids had been eliminated except for a thin line, probably present but not fossilized, that would lead to the modern stomatopods. All of the late Paleozoic primitive malacostracans (the belotelsonids, waterstonellideans, eocaridaceans, the early palaeocaridacean syncarids, the pygocephalomorphan mysidaceans) were now absent. The sessile epibenthic suspension feeding community (many of the bryozoans and brachiopods, many primitive echinoderms such as blastoids and most of the crinoids, primitive tabulate, and rugose corals) that had dominated Paleozoic benthic habitats had been largely obliterated. Most of the plankton was gone. As the Triassic progressed, however, mammal-like reptiles and the early groups of dinosaurs appeared, other reptiles diversified, and coniferous trees came to dominate the land. In the sea, the surviving remnants of the cephalopods began to diversify into modern groups, including predaceous squid- and octopus-like forms. Scleractinian corals began to proliferate, even though they were subordinate to sponges, algae, and bryozoans as reef builders. Conchostracan branchiopods, ostracods and lepadomorph cirripeds, anaspidacean syncarids, lophogastrid mysidaceans, tanaid and isopod peracarids, the first dendrobranchiate decapod shrimps, and astacid and palinurid decapods are recorded. Probably many more lineages were reestablishing themselves, though unfossilized, during this period (remipedes, cephalocarids, all of the branchiopod groups, many of the primitive maxillopods, other primitive eumalacostracans such as syncarids, and the mysidaceans, peracarids, and eucarids).

Another mass extinction, however, spread biological devastation in the Late Triassic.

It is thought that huge continental landmasses begin to accumulate heat beneath them, creating convection currents in the mantle that cause landmasses to drift apart approximately every 500 million years. In the Jurassic (210–140 mya), Pangaea began to break apart. About 200 mya, Laurasia (North America and Eurasia) moved away from Gondwanaland, forming a nascent Atlantic Ocean in the west and a gulf between Asia and East Africa in the east. India drifted away from Antarctica toward Asia, and the south Indian Ocean began to form between Australia (still attached to Antarctica), Antarctica, and the joined southern tips of Africa and South America. A widening rift between South America and Africa was flooded with seawater near the end of the Jurassic. The northern drift of Laurasia created a pathway for a circumtropical current between the northern and southern continents. This shallow warm seaway, adorned with archipelagos, would be called the Tethys Sea, and it was the cradle from which most of the modern families of organisms that we know today would emerge.

The early Jurassic climate was cool, becoming warmer. High sea levels encroached on the continents, spreading shallow seas. Although mass extinctions challenged the faunas during the Middle and Late Jurassic, the Jurassic period is known for its spectacular radiation of reptiles, including the dinosaurs and pterosaurs. Modern groups of amphibians (frogs, salamanders) and birds arose. Primitive flowering plants and many of the modern groups of insects (flies, moths, bees, wasps, ants) diversified in the increasingly warm, mild climate. In the sea, marine reptiles must have been fearsome predators, crocodiles and turtles abounded, many of the modern families of boney fish were formed, corals acquired symbiotic algae that allowed them to grow up and away from sedimentation (such as that created by burrowing animals in the late Paleozoic) 10 times faster than they could in the absence of the photosynthetic symbiont. Reefs became more extensive and boring organisms more common. The Jurassic fossil record of Crustacea includes conchostracan and notostracan (extinct Kazacharthrans) branchiopods, and shows a renewed ostracod radiation. The lepadomorph barnacles were present, and the balanomorph (acorn) barnacles made their first appearance. Many of the other groups (remipedes, cephalocarids, other branchiopods, other maxillopods) likely were present but not fossilized. In the Malacostraca, the fossil sculdid hoplocarids (descendents of the Devonian archaeostomatopods and progenitors of the modern stomatopods) arose. The last remaining vestige of the phyllocarids (the leptostracans) and some syncarids may have been present but not fossilized. The first mysidacean eumalacostracans appeared, tanaid and isopod peracarids were present, and many of the other peracarids likely were present but not preserved. The eucarid malacostracans burst on the scene, as shown by the first appearances of fossil caridean, stenopodidean, and thalasinidean shrimps, and anomuran and definitive brachyuran crabs. Dendrobranchiate shrimps, as well as astacidean and palinuran crayfishes and lobsters, also were present in the fossil record, completing the roster of all the major groups of eucarids known today. Despite the mass extinctions, no major lineages were lost during the transition from the Jurassic to the Cretaceous.

The tropical Tethyan seaway continued to girdle the earth and maintain mild conditions during the Cretaceous (140–65 mya), although the pathway was narrowing in several sites (e.g., northeastern and northwestern Africa). North America was still united with Eurasia despite a widening gulf invading northward from the opening Atlantic. India had not yet crashed into Eurasia, Madagascar was sliding northeastward from southern Africa, and Australia had begun to separate from Antarctica, beginning its trek toward the tropics. Africa and South America had left Antarctica behind as they migrated northward and severed their connections to each other. The Andes and Rocky mountains rose as North and South America ploughed westward and the Pacific plates dived beneath them. The continents were awash with spreading inland seas and swamps. The reptilian radiation was at its apex, spawning the largest land and marine reptiles ever known. Grasses appeared, and other flowering plants and insects diversified even further; coniferous gymnosperms declined.

In the sea, reefs began to accrete and proliferate on an unprecedented worldwide scale and reached their highest diversity ever. Scleractinian corals flourished, and in sheltered areas huge rudist bivalve mollusks (which, like the corals, may have obtained photosynthetic symbionts) committed massive deposits of calcium carbonate to reef structure, threatening the dominance of corals in the reef community. The teleostean (boney) fishes underwent major diversification, developing new body shapes that increased their maneuverability, calcareous mouth plates that crushed shelled and hard-bodied prey, and pincer-like jaws that could grab and pluck prey from the substratum. Benthic octopuses with flexible arms and bodies could engulf and subdue prey. Crustaceans developed hardened specialized claws that could crush or cut prey. Motile and sessile organisms seeking refuge from the burrowing organisms that caused extensive movement of benthic sediment (bioturbators) likely colonized the high, hard substratum provided by the corals and rudists. Boring organisms proliferated worldwide, for boring into the calcareous substratum provided protection from the increasingly diverse and dangerous predators. Individual bioeroders eventually died, however, leaving a vacant crypt. This three-dimensional microenvironment provided a safe refuge for a multitude of small sessile and motile invertebrates. The organisms that invaded the holes left by the borers, termed the cryptofauna, were constrained to small body sizes in order to occupy the protective holes, were plagued by predation whenever they emerged from the crevice to feed or mate, and had to develop special adaptations for acquiring and keeping the packed refuge space. All of these processes accelerated morphological and other adaptations in the benthic fauna. Small body size usually is associated with production of fewer, larger eggs, abbreviated development, and higher rates of speciation and extinction in motile invertebrates. These life history features likely facilitated diversification of other behavioral and morphological features, such as the deadly fighting behavior and armored bodies and appendages that developed in many of the stomatopods and decapods. Special adaptations (including coloniality, unique growth strategies, and chemical warfare) also developed in the sessile cryptofauna, promoting diversification and enabling them to better acquire and defend living space.

Among the crustaceans, the conchostracan branchiopods and ostracod maxillopods again are represented in the Cretaceous fossil record; the remipedes and all other modern groups of branchiopods and maxillopods likely were present but not preserved. Copepods, ascothoracidan barnacles (as trace fossils), verrucomorph (wart) barnacles, and the squat modern cthamaloid acorn barnacles made their first appearance as fossils. Among the lower malacostracans, the leptostracan phyllocarids likely were present but unfossilized, and the predaceous stomatopod hoplocarids came into their own. The extinct sculdids (bathysquilloids) and three of the major modern groups of stomatopods (the reef-dwelling gonodactyloids, mud-burrowing squilloids, and sand-burrowing lysiosquilloids) all appeared for the first time. The syncarids and mysidaceans probably were present but not recorded. Although their distribution suggests a more ancient derivation, the thermosbaenaceans are first known in the Cretaceous. Other peracarids and other eucarid shrimps (stenopodideans, thalassinideans) probably were present, but the tanaidaceans and isopods, as well as the dendrobranchiate and caridean shrimps, definitively inhabited Cretaceous seas. The advanced eucarids (astacids, palinurans, anomurans, brachyurans) were all present and continued their radiations begun in the Jurassic.

Following a mild mid-Cretaceous mass extinction, a significant mass extinction at the end of the Cretaceous eliminated the dinosaurs and many of the great reptiles (including the huge flying and marine reptiles), the ammonoids and many of the older cephalopod lineages, the reef-building rudist bivalves, and many groups of plankton. It is likely that the explosive impact of a giant asteroid contributed to this biotic disaster. The role of additional factors in the faunal turnover between the Mesozoic and the Cenozoic is debated. Mass extinctions often determine the biotic structure of the next era by filtering what major groups survive to radiate again in the postextinction periods. This was true for the mammals and birds, relatively minor groups in the Mesozoic that persisted through the mass extinction and then, in the absence of the great reptiles, diversified and dominated the ecology of terrestrial environments in the Cenozoic. For the crustaceans, however, the biotic upheaval at the end of the Mesozoic had far less impact than had the great extinction at the end of the Paleozoic. The kazacharthran branchiopods, related to the notostracans, had been lost during the Jurassic, but the lineage continued and survives today. The hoplocarids also persevered. Despite the possible loss of the sculdid stomatopods at the end of the Cretaceous, the bathysquilloid lineage persisted (though restricted today to deep continental shelf environments). The new stomatopod superfamilies generated in the Cretaceous endured and are common in warm shallow waters today. No other major crustacean lineages were lost. Their small body sizes (compared to those of some of the giant reptiles and cephalopods, for example) may have allowed crustaceans to persist in a time of great ecosystem disturbance and food limitation. Also, although the extent of the geographic range of an individual species confers resistance to background extinction, a broad distribution for the entire group is one of the most important factors enabling lineages to survive mass extinctions (Jablonski, 1991). Reaping the success of radiations in the warm, reef-studded seas of the Jurassic and Cretaceous, many crustacean groups likely had worldwide distributions that enabled them to survive the end of Mesozoic mass extinction.

The Cenozoic can be divided into the Tertiary (65–1.6 mya, including the Paleocene, Eocene, Oligocene, Miocene, Pliocene) and the Quaternary (1.6 mya–present, including the Pleistocene and Holocene). The Atlantic was now formed and widening. India crashed into Eurasia, creating the Himalayan mountains. Extensive folding, resulting in the Alps, Caucasas, and other mountains, occurred as parts of Africa fused with Eurasia, forming Asia Minor and constricting the ancient Tethys seaway into the isolated, sometimes dry, Mediterranean and Black Sea areas. Mountain building continued in the Rockies, the Coast Range, and the Andes, eventually shifting the flow of the Amazon from westward to eastward. The early Tertiary climate was warm, wet, and mild. The great radiations of placental mammals began. Horses, camels, rhinoceroses, dogs, cats, whales, bats, and primates diversified in Eocene terrestrial environments. A gradual cooling trend began in the Oligocene. Extensive plains, grasslands, and new mammalian lineages (deer, elephants, antelopes, giraffes, and the ancestors of cattle and pigs) developed through the Miocene, and large mammalian carnivores radiated in the Miocene and Pliocene.

In the oceans, there were no extensive coral reefs in the Paleocene, although scleractinian corals had survived the mass extinction at the end of the Mesozoic. The Eocene brought a reradiation of reef-building corals and many families and genera of fish, mollusks, and crustaceans that survive today. Extensive volcanism and tectonic movements in the first half of the Tertiary set the stage for island fringing reefs and the development of atolls in the Pacific.

The Tertiary produced a new radiation of the cirriped barnacles, including the coronuloids (which live as commensals on whales and turtles and appeared in the Paleocene) and the first modern intertidal balanoids (which originated in the Eocene and diversified again in the Miocene). The ostracods radiated once again during the Eocene; they also are represented in the Paleocene and Miocene fossil records. Among the branchiopods, the first definitive anostracans and cladocerans appeared during the Eocene. In the stomatopod hoplocarids, a possible sculdid bathysquilloid, the squilloids, and the first pseudosquilloids are known from the Eocene; in addition, fossil gonodactyloids and squilloids are recorded from the Miocene, and squilloids from the Pliocene. Following a late Paleozoic genesis, the isopod peracarid eumalacostracans radiated during the Eocene, and the first amphipod peracarids appeared; both of these groups are known also from the Oligocene and Miocene. Among the eucarid malacostracans, the caridean shrimps left fossils in the Oligocene and Miocene, and the thalasinidean shrimps are recorded from the Miocene; both groups hailed from the Jurassic. The astacid crayfish and lobsters, among the first of the Paleozoic decapods, are known from the Eocene. The brachyuran crabs, also from a likely Paleozoic heritage, underwent major radiations in both the Eocene and Pliocene, with many of the modern freshwater and terrestrial lineages developing in the latter epoch. Brachyurans are represented in the Paleocene and Miocene fossil record as well. Anomurans are known from the Oligocene and Miocene.

One of the most significant events in the late Tertiary was the gradual establishment of a complete barrier to circumtropical ocean circulation about 3 mya as mountain building along the west coasts of North and South America created a Central American isthmus that linked the two continents. This removed the ameliorating effect of the circumtropical current on the climate and created a much smaller, basin-like West Atlantic region as well as a long thin East Pacific region. Considerable extinction and divergence of new genera and species ensued in the New World tropical seas.

In spite of a minor mass extinction at the end of the Eocene and deteriorating ocean circulation in the Miocene and Pliocene, however, no major groups of crustaceans were lost during the Tertiary. The ostracod and cirriped maxillopods, the stomatopods, the isopod and amphipod perarcarids, and the brachyuran decapods all underwent major radiations, particularly during the Eocene.

The Quaternary (Pleistocene 1.6–0.01 mya, Holocene 10,000 years ago-present) ushered in four major periods of glaciation, particularly on the northern continents (North America, Europe, North Asia) that had not been extensively glaciated during the Paleozoic and Mesozoic. Early humans appeared in the Pleistocene. Mountain building in the Cascades, Andes, Caucasus, and the Himalayan regions continued, and rifts in eastern Africa, Asia Minor, Baikal, and southeast Asia reflected continuing tectonic shifts. Worldwide sea levels receded with the advance of continental glaciers and advanced during warmer interglacial intervals. The current climate represents emergence from a colder glaciated interval about 10,000 years ago. The extent to which the frigid polar glaciation impacted climate and seawater temperatures at lower latitudes is debated. Most modern coral reefs (including Australia's Great Barrier Reef) are only about 10,000 to 15,000 years old and represent a thin surface veneer on a complex topography inherited from past reef communities as they adjusted to changing sea levels. The anostracan branchiopods, isopod and amphipod peracarids, and brachyuran crabs are among the relatively few crustacean groups that were preserved as fossils during the Pleistocene. Although morphological and distributional evidence suggests that some of the smaller and thinly skeletonized lineages were present much earlier, many groups are recorded for the first time during the Holocene. These include the nectiopodan remipedes, cephalocarids, mystacocarids, branchiurans, tantulocarids, ascothoracidan and rhizocephalan cirripeds, bathynellacean syncarids, thermosbaenaceans, mictaceans, euphausiaceans, and amphionidaceans. Many other groups have radiated or renewed their diversification during the Holocene. Showing remarkable evolutionary resilience since the beginning of the Paleozoic, the ostracods are undergoing a fifth major radiation. The cirripeds show considerable current diversity. The stomatopods, peracarids (especially isopods and amphipods), and decapods (dendrobranchiate shrimps, and especially the caridean shrimps, thalassinidean shrimps, anomurans, and brachyurans) all are undergoing major modern diversification. Although no major lineages have been lost during the Quaternary, today the remipedes, cephalocarids, branchiopods, remaining lower malacostracan lineages (phyllocarids, syncarids) and some of the ancient eumalacostracan lineages such as the palinurans and astacids often inhabit peripheral refugia and appear to have their greatest eras behind them.

C. Ecological Diversity in Crustaceans

1. The Habitats That Crustaceans Occupy

The major types of environments inhabited by crustaceans can be separated into 16 major, often overlapping, categories. Of all of these, crustaceans are most commonly found in benthic marine habitats, including both level bottom and coarse (rock and reef) habitats (35 groups). Brackish water (coastal deltas and lagoons, brackish marine caves and tidepools, salty inland pools, and brackish groundwater) is the next most common habitat for which 23 groups (59%) of crustaceans are well suited. Seventeen taxa (44%) are pelagic in the upper part of the ocean (<200 m), and 18 (46%) groups are found in the deep sea (200 to >6000 m), either as pelagic or as benthic species on level bottoms or around hydrothermal vents. Fifteen taxa (38%) inhabit permanent bodies of fresh water (ponds, lakes, streams, rivers), including both benthic and pelagic components. Eight groups (21%) contain species that live in intertidal environments (some of these among sand grains), and 8 (21%) inhabit other semiterrestrial or terrestrial conditions (usually moist conditions or decaying vegetation). Some isopods, some amphipods, and a few brachyurans are truly terrestrial, having achieved independence from water for reproduction. Eight taxa (including 5 highly specialized groups) have adopted parasitism (21%), and 7 groups (18%) include commensals that live on or in other organisms. Seven groups or sections of taxa (18%) flourish at high latitudes, 5 (13%) inhabit either marine or freshwater caves (with various degrees of specialization), and 5 (13%) earn a fragile existence in inland ephemeral freshwater and briney pools. Three taxa (8%) can tolerate hypersaline (very salty) conditions, 3 (8%) are known from hot springs, and 3 (8%) live a specialized existence in groundwater. At least 1 group lives interstitially among the meiofauna (organisms smaller than 0.5 mm) in subtidal benthic marine sediments. The number of major taxa that occupy each of the various types of habitat may but does not necessarily reflect the number of species in that environment. This analysis, instead, demonstrates the diversity of major adaptations (in morphology, physiology, and behavior) that crustaceans have developed to accommodate particular environmental challenges. These adaptations provide the underlying framework, overlain by other specializations, on which evolution can proceed in the long run.

The same environmental categories can be used to examine ecological diversity among the different crustacean groups. It is likely that those groups that occupy the greatest diversity of habitats are the ones that will be able to survive massive environmental change, since not all habitats are likely to be equally affected by most forms of environmental disturbance. As shown in the discussion of taxonomic diversity, the maxillopods and the eumalacostracans (particularly the peracarids and eucarids) are the champions in terms of ability to occupy a diversity of different environments. Following a history of evolutionary success that stretches from the earliest Paleozoic to the present, the ostracods inhabit seven different types of environments. The comparatively recent (late Mesozoic) but ubiquitous copepods also occupy seven major types of habitats, including exceedingly specialized parasitic ways of life, and the ancient thoracican barnacles live in five major habitat types. Among the lower malacostracans, the anaspidaceans are surprising in the diversity of seemingly marginal types of environments that they inhabit (five: permanent fresh water, ephemeral freshwater pools, caves, semiterrestrial, groundwater), and the mysidans have radiated into seven major habitats and lifestyles. Even more remarkable, the thermosbaenacean peracarids, with only 11 species, inhabit five different types of environment as categorized earliear (benthic marine, inland salt ponds, fresh and salty groundwater, caves, hot springs). The tanaid peracarids and the anomuran eucarids occupy five and six major types of habitats. The caridean shrimps and brachyuran crabs (decapod eucarids) and the isopod and amphipod peracarids, however, inhabit more different types of environment than any other groups of crustaceans (seven, eight, nine, and nine types of habitat, respectively). Although many carideans and some brachyurans live commensally with other organisms, a fully parasitic existence is the only habitat that these eucarids have not exploited as fully as have some isopods and amphipods. Although caridean shrimps have been at least as successful in fresh water and are more diverse in pelagic environments than the brachyurans, they have failed to invade terrestrial habitats as completely as have the brachyurans, amphipods, and isopods. The isopods are the most successful terrestrial crustaceans, and they, along with most other peracarids, flourish in cold, high-latitude environments as well as the deep sea. Amphipods are less specialized for parasitism and terrestrial existence than the isopods. Like the amphipods and isopods, the carideans and brachyurans have radiated into all major marine environments, including the deep sea and its hydrothermal vents. The carideans likely invaded fresh water very early (Mesozoic), whereas the brachyurans undertook major invasions of freshwater and terrestrial habitats in the Late Tertiary and Quaternary.

A number of crustaceans also modify their microenvironment by commandeering empty shells or other structures, engineering tubes or burrows, and camouflaging themselves from pervasive predation in various ways. Anomurans, tanaids, and amphipods all inhabit vacated gastropod shells; anomuran hermit crabs also inhabit tusk shells, hollow twigs and bamboo, calcareous worm and vermetid mollusk tubes, hollow pieces of coral, and even bottle caps. Some anomurans plant anemones that function in camouflage and defense on their gastropod home. Some dromiid brachyuran crabs carry a clam shell on their backs, and lithodid and majid brachyurans culture domes of sponge or decorate their carapace with algae. Tanaid and amphipod peracarids construct pebble tubes, and amphipods modify algal structures for domiciles or camouflage. One group of amphipods bores into wood, one group of isopods bores burrows in silt and clay, and acrothoracican barnacles bore into coral to make protective lodgings. Coral-dwelling stomatopods chip, scrape, and modify preexisting holes in coral or reef substratum to suit their needs, and most groups of crustaceans are well represented in preformed crevices and bioeroded holes in hard, especially calcareous, substrata (coral, limestone and reef basement, oyster and tube worm reefs). Many of these organisms, such as the thalassinideans, line and modify the preexisting crevice with silt and mucous for a better fit. Some groups of caridean snapping shrimps construct smoothly lined networks of burrows within sponges. Large species, such as stone crabs, coral crabs, spider crabs, spiney lobsters, and some stomatopods inhabit refuge space beneath bioeroded ledges and caverns on reefs. Several groups of stomatopods, astacid crayfishes and lobsters, thalassinidean shrimps, and brachyuran land crabs, fiddler crabs, and ghost crabs all excavate large (in some species very complex) burrows in mud and sand environments. All of these habitat modifications reflect the pervasive intensity of predation by epibenthic and benthic predators such as cepahlopods (particularly octopuses), other crustaceans, and especially fishes. By increasing the complexity of the microenvironment, many of these structures increase the potential for commensal associations and heighten the capacity of the environment to support high species diversity.

2. The Role of Crustaceans in Biological Communities

Examination of the trophic position of crustacean groups indicates that most crustaceans have relatively generalized feeding habits and most are subject to strong predation. Their abundance, moderate size, and generalized carnivory likely exerts top-down control in some communities. Also, as evidenced by almost universal morphological and behavioral adapatations against predation as well as their importance in the diets of many different predators, the availability of crustaceans as prey for other trophic levels probably is essential for the structure, function, and diversity of many aquatic communities. Cladocerans and copepods support higher trophic levels, including fishes, in freshwater habitats. When their ephemeral habitats are located, anostracans and notostracans are heavily consumed by birds. Anaspidacean syncarids are heavily preyed upon by freshwater fishes if the shrimp are dislodged from their refuges. Stomatopods, mysidans, tanaids, cumaceans, isopods, amphipods, dendrobranchiate shrimps, caridean shrimps, anomurans, and brachyurans all are common in both epibenthic and benthic fish stomachs; cumaceans, isopods, and amphipods are particularly important in the deep sea. Dendrobranchiate shrimps and mud-dwelling stomatopods are important in the diets of many large, commercially important fish, and euphausiaceans and pelagic anomurans are a major source of food for whales, squid, tunas, and seabirds.

Most crustaceans are either generalized predators and scavengers (30 taxa or portions of taxa), or detritus and suspension feeders (35 taxa or parts of groups), Many biologists consider the predaceous, scavenging remipedes to exemplify the primitive feeding mode, but detrital suspension feeding, as seen in the phyllopodous cephalocarids, branchiopods, and some of the primitive phyllocarid malacostracans, obviously can be traced far back in crustacean history as well. Herbivory is not particularly well developed in crustaceans, although many of the phyllopodous groups (cephalocarids, branchiopods, leptostracans) may ingest phytoplankton and plant material along with other detritus, The suspension feeding ascothoracidan, acrothoracican, and thoracican barnacles consume phytoplankton; and some ostracods, isopods, amphipods, and brachyurans feed on micro- or macroscopic plants (some of these species also eat detritus or other organisms as well). Euphausiaceans, however, are important grazers of phytoplankton at high latitudes, and copepods may consume more primary productivity than any other organism on earth because of their ubiquity and consumption of phytoplankton in the largest habitat on earth, the open ocean. The mouthparts of crustaceans are relatively easily modified for piercing and tearing flesh, so strict carnivory and ectoparasitic habits, with the potential for considerable impact on host populations, have evolved in a number of groups, including a cladoceran, branchiurans, tantulocaridans, some ascothoracidan barnacles, isopods, and amphipods. Parasitism is most extremely developed in the endoparasitic copepods and rhizocephalan barnacles; many ascothoracidan barnacles and some isopods also live inside and consume the energy or tissues of their hosts. Large body size and armored claws that smash, crush, or shear hard-bodied prey make stomatopods, palinuran and astacid lobsters and crayfish, and anomuran and brachyuran crabs potent predators that, near the top of their benthic food webs, exert important effect on their local communities.

3. Life History Patterns of Crustaceans

Although most crustacean groups reproduce sexually, a number of groups, particularly in fresh water and high latitude environments, produce offspring without exchanging eggs and sperm. "Parthenogenesis," a form of asexual reproduction where eggs undergo development without fertilization, is found almost exclusively in the Branchiopoda and Maxillopoda. Other forms of asexual reproduction, such as budding of new individuals from a previous individual, do not appear to be present in the Crustacea, although regeneration of missing body parts is common. In the branchiopods, most anostracans reproduce sexually (and produce fertilized resistant eggs that are dormant during unfavorable conditions), but some populations are exclusively parthenogenetic. Parthenogenesis is common in both notostracans and cladocerans. Some notostracan populations are exclusively parthenogenetic, especially at high latitudes. Notostracans also reproduce sexually, producing fertilized resistant eggs that can withstand harsh conditions. Some species of cladocerans produce males and undergo sexual reproduction only during periods of environmental deterioration, such as autumn; males may be dwarfed and short lived. Some species of conchostracans are exclusively parthenogenetic, and other species may include both parthenogenetic and sexual individuals. In the maxillopods, some freshwater ostracods are parthenogenetic, and other species are both parthenogenetic and sexual. Some harpacticoid copepods are parthenogenetic. Some copepods produce resting eggs that are dormant for extended periods. Of the rare deep-sea tantulocarids, only females are known, though parthenogenesis is not proven. Evidence suggests parthenogenesis in some thoracican barnacles. Some terrestrial isopods are parthenogenetic.

Reduction of the prominence of males, as evidenced by short-lived, usually dwarfed males that frequently live on or in the female, is seen in the freshwater cladocerans, parasitic ascothoracidan barnacles (dwarf male free living or lives on female), boring acrothoracican barnacles (lives on female), sessile thoracican barnacles (lives on hermaphroditic individual or female), parasitic rhizocephalan barnacles (short-lived dwarf male invades extruded female gonad), and parasitic isopods (lives on female). In the deep sea gnathiid isopods, the second maxillipeds are reduced to flaps that cover the mouth of the nonfeeding male. Few males are known (although sexual dimorphism appears in development) in the pelagic deep sea amphionidaceans, and populations of the primitive procaridid shrimps are composed mostly of females; these skewed sex ratios may be the result of differential mortality in males, reduction in the prominence of the male life cycle, or flexible sex determination (as is common in many crustacean groups) rather than a propensity for parthenogenesis.

The reproductive organs of crustaceans are relatively simple, and both testes and ovaries apparently develop readily, with dominance shifted from one to the other depending on hormonal and environmental conditions. The primitive remipedes and cephalocarids are hermaphroditic. Although most sexually reproducing branchiopods are gonochoristic, some species of notostracans are hermaphroditic and likely undergo sex reversal. In the maxillopods, the sex of the sessile cirripeds is notably flexible. Although most of the parasitic ascothoracidan barnacles are gonochoristic, species in one group are simultaneous hermaphrodites and those in another group are protandrous hermaphrodites. In others, the first arrival to a host becomes female, and the next arrival becomes a male. Most of the sessile thoracican barnacles are simultaneously hermaphroditic or mixed hermaphroditic and separate sexes (dwarf males may reside either on females or hermaphrodites). Most acrothoracican and rhizocephalan barnacles, however, are gonochoristic. The other center of sex change is found in the malacostracans. In the peracarids, tanaidaceans usually are protogynous hermaphrodites, or the same population may have both gonochoristic and hermaphroditic individuals. Sex change appears to be influenced by environmental factors, possibly even the ratio of males to females as affected by predation. Most isopods are gonochoristic, but some (especially freshwater) lineages change sex. Both protandrous and protogynous species are known, and both gonochoristic and hermaphroditic individuals may occur in the same population. Particularly in parasitic isopods, sex is determined by the presence or absence of another individual; first arrivals become females and the second arrival becomes a dwarf male that resides on the resident female. Most amphipods also are gonochoristic, but protandrous species are known, and some species are composed of both protandrous and gonochoristic individuals. Although most caridean shrimps have permanent separate sexes, some species are simultaneous hermaphrodites and some are protandrous. Both gonochoristic and hermaphroditic individuals may occur within some populations. Sexual differentiation may be affected by both social and environmental factors. For example, some species of snapping shrimp live in colonies of up to 350 individuals in a sponge, with only one large reproductive female, the "queen." A host of smaller male or nonreproductive "workers" defend the colony and maintain communal burrows.

Sexual reproduction often is accompanied by morphological structures that have differentiated between males and females, probably as the result of sexual selection. Males often are adapted for greater motility than females. In the ascothoracidan barnacles, the female's body is enveloped entirely within the carapace, while most of the swimming appendages are exposed in males. Male tanaids have well-developed pleopods, but abdominal appendages are absent in females; males also have larger eyes and more antennal sensory structures than females. Male cumaceans bear abdominal pleopods while females do not, the antennae are reduced in females but not males, and the eyes are more reduced in females than males. The carapaces of female thermosbaenaceans and amphionidaceans are very enlarged, while those of the males are reduced, exposing swimming appendages. The last pair of thoracic legs are absent in female amphionidaceans but not in males, and the females, with reduced mouthparts and a vestigial gut, do not feed.

Various appendages have been modified in many groups to facilitate mating. The first pair of maxillae (maxillules) were modified as a clasper in males of the Devonian Lipostracans (allied to the cephalocarids). The antennae are modified for detecting or holding females in male anostracans; some cladocerans, ostracods, and cumaceans; and in many amphipods. The size and shape of the antennules of male leptostracans and euphausiaceans and of some species of copepods are modified, probably for clasping females. Other secondary sexual characteristics include copulatory structures, which are common in crustaceans. Anostracans have a pair of eversible penes on the first two abdominal segments. The leaflike appendages of the first two trunk segments are modified for copulation in conchostracans; the first two trunk limbs of some genera of ostracods are modified to assist mating, and the third trunk appendages often are modified into copulatory organs. The fourth trunk appendages, sixth thoracopods, and second through fourth thoracopods of male mystacocarids, copepods, and branchiurans, respectively, are modified for copulation. Abdominal appendages have been fused into a single long median penis, the only remaining vestige of the abdomen, in thoracican barnacles. Paired penes also are fused into a single median penis in the tantulocarids and some isopods. A pair of copulatory tubes are associated with the eighth thoracopods in male stomatopods, and the eighth thoracopods are modified, probably for copulation, in the bathynellaceans. The bases of the eighth thoracopods in mate thermosbaenaceans bear a pair of well-developed penes. Male mictaceans sport a copulatory process on the eighth abdominal segment, and, although pleopods are otherwise absent in both males and females, male mictaceans retain the second pair of pleopods, probably to assist transfer of sperm. The anterior abdominal pleopods are structurally modified to assist copulation in many male malacostracans, including stomatopods, mysidans, isopods, amphipods, and dendrobranchiate shrimps. The anterior pleopods of both male and female euphausiaceans are elaborately modified for copulation.

Several groups of crustaceans exhibit strong sexual dimorphism in body size and other structures, including the late Paleozoic pygocephalomorph mysidaceans. Sexual dimorphism can be observed among developing juvenile amphionidaceans after stage 9. Male tanaids show pronounced sexual dimorphism in a variety of sensory structures as well as relative size of the chelipeds. The overall morphology of male and female tanaids becomes so highly modified as they molt into copulatory versus noncopulatory phases that they frequently have been misclassified as different taxa. Amphipods and caridean shrimps show similar variability of body size and shape between copulatory and noncopulatory stages, probably reflecting the impact of strong selection for attracting and defending mates against other suitors. Some groups of isopods (especially terrestrial lineages) show extreme dimorphism in size, with males greatly exceeding the size of females. The claws of many decapods (including some caridean shrimps, some anomurans, and particularly some thalassinidean shrimps, astacid crayfish and lobsters, and many groups of brachyuran crabs) are greatly enlarged in males. Either both or only one of the claws may be enlarged for fighting or behavioral displays. Mating or territorial displays often are associated with enlargement of only one claw, as in the fiddler crabs that attempt to attract females to their burrows with colorful choreographic waving displays. Though not sexually dimorphic, stomatopods also seek to attract mates by exposing species-specific colored eyespots and irridescent patterns on their spread maxillipods. Several groups, including land crabs and fiddler crabs, use sound vibrations (usually stridulating or thumping the walls of their burrow) to attract mates. Various other crustaceans stridulate (stomatopods, some isopods, palinurans, other brachyurans), but these sounds may function to deter attacking predators rather than to attract mates. Bioluminescent photophores on euphausiaceans may facilitate sexual identification, although this is not proven. One of the most sensational courtship displays in crustaceans is created by male ostracods that emerge from coral reef substratum about an hour after dark, emitting sequential puffs of bioluminescent material from the maxillary glands in a species-specific pattern as they ascend toward the surface. Different species deposit the discrete puffs of light in uniformly spaced vertical strings, suspend them in a line at a 45° angle from the substratum, or place sequential clusters at different heights. Females do not emit bioluminescent puffs, but emerge from the reef and follow the trail of lights to mate with the displaying males. A horde of nonluminescent males also follow the light trails, apparently to "sneak" matings with the pursuing females.

Most crustaceans produce round nonmotile sperm, although the sperm of some maxillopod groups are flagellated. The sperm, released either in seminal fluid or in packets of spermatophores, are deposited directly into the female genital openings or into a fused medial receptacle. Females may store sperm for long periods (up to several years in some astacid lobsters), although subsequent matings can dilute the sperm of a particular male.

Most crustacean matings are relatively brief (minutes to hours), but some crustaceans undertake long arduous courtships and maintain long-term pair bonds. Since little paternal care or mate provisioning is documented, most pair bonds probably function to ensure that another male does not also fertilize the female. Males of isopods, amphipods, and brachyuran crabs often sequester females, holding and carrying the smaller female for extended periods prior to copulation, which frequently follows female molting. In the violently aggressive and solitary stomatopods, mating is accompanied by an extended period of courtship (involving both visual displays and tactile stroking) that gradually mediates the tendency to smash the potential mate with the clublike or spearing raptorial appendages. The male and female reside together (either in the female's or male's burrow, depending on the taxonomic group) for a number of days, copulating repeatedly. After as much as a week, the female's aggressive levels rise and she evicts the male before laying the eggs. Commensal pinnotherid crabs ("pea crabs," which inhabit oysters and other organisms' burrows), parasitic isopods, stenopodid shrimps, and many groups of caridean shrimps usually are found in pairs, and probably maintain extended (seasonal or permanent) pair bonds. Some deep-sea stenopodid shrimps apparently colonize glass sponges as postlarvae, and the male and female (perhaps the opposite sex is induced in the later arrival) become permanently imprisoned in the crystalline gridwork of the sponge as they grow. These life-long pair bonds are the source of a romantic oriental legend.

Most crustaceans brood their young, but there is a surprising diversity in methods of treating the young, and some release the embryos free into the environment. The remipedes probably shed their embryos directly into the water (no brooding structures have been observed). In the branchiopods, the notostracans brood their embryos very briefly, then deposit them on the substratum, while most cladocerans release their embryos into the water. In the maxillopods, some ostracods release embryos directly into the water and some deposit them on the substratum. The interstitial mystacocarids apparently lay their embryos free among sand grains. Some copepods release embryos directly into water. Branchiurans leave their host to attach their embryos to a plant or rock. In the malacostracans, the extinct Late Paleozoic belotelsonideans, waterstonellideans, and palaeocaridaceans probably shed their embryos directly into the water, since no structural adaptations for brooding have been seen. Contemporary anaspidacean syncarids, and probably bathynellaceans, release their embryos directly into the environment. Contemporary stomatopods deposit their eggs in a free wheel-shaped disk that the female holds in her maxillipeds, grooming and defending them for several weeks. Most euphausiaceans and most dendrobranchiate shrimps shed their embryos directly into the water.

However, most crustaceans exhibit a variety of structural, physiological, and behavioral modifications to better protect their progeny. Cephalocarids brood one or two embryos at a time on anterior abdominal processes, and the allied Devonian lipostracans had a flaplike egg pouch. Among the branchiopods, the anostracans carry their eggs briefly in an egg sac before releasing them. Some cladocerans shelter their eggs in their carapace; the embryos even obtain nutrients from the sides of the brood chamber in some species. One species of cladoceran broods eggs in the molted carapace. Conchostracans carry their embryos on the thoracopods inside the bivalved carapace. Most ostracods brood their young inside the bivalved carapace. Copepods usually drag an egg sac behind the body, but some brood embryos on the sixth thoracopods. In the ascothoracidan cirripeds, the female's carapace is enlarged (enclosing her body) for brooding the young. Embryos are released at the nauplius stage or retained in the carapace through all larval stages up to the cyprid. The embryos of acrothoracican cirripeds hatch at an early stage but are brooded in the carapace through the cyprid stage. The body of thoracican barnacles is engulfed by a fleshy mantle cavity in which the embryos are brooded through the nauplius stage. In the malacostracans, leptostracan females brood embryos under their carapace until after two larval molts. As noted earlier, female stomatopods undertake extensive care and protection for their free egg mass within their cavities in coral or excavated burrows in sand or mud. Female mysidans bear a brood pouch on the last two or three thoracopods. The enlarged, thin, blood-filled carapace of female thermosbaenaceans serves as a brood chamber as well as for respiration. Similarly, the carapace of female amphionidaceans is thin and greatly enlarged, and it may be used for brooding. In this group, the first pair of pleopods, greatly enlarged and extending anteriorly beneath the carapace, are thought to assist in brooding the young. The brood pouch is particularly well developed in the tanaids, cumaceans, mictaceans, isopods, and amphipods. This "marsupius" is formed from expanded flaps, called oostegites, at the bases of thoracopods 3 to 6 in tanaids and cumaceans. Sometimes the oostegites are restricted to thorapod six in tanaids. Cumaceans, like isopods, bear an additional oostegite projecting toward the posterior from the maxillipeds. Female mictaceans bear oostegites either on thoracopods 2 to 6 or 3 to 7. Female isopods have oostegites on the second to fifth, second to sixth, or all eight thoracopods. One family forms a brood pouch from ventral extensions of the body wall rather than oostegite projections from the base of the legs. Several types of internal brooding are present in isopods, including paired invaginated pockets in the cuticle with narrow openings and a single median pocket extending into the thorax between the sixth and seventh segments. Two groups (the gnathiids and flabelliferans) brood young within the body cavity itself; in the flabelliferans, the enlarged oviducts serve as uteri and the oostegites are reduced to small flaps covering the gonopore. Female amphipods develop oostegites on the third to fifth, third to sixth, or fourth to fifth thoracopods; many amphipod taxa also have simple or elaborate gills that protrude from the base of the thoracopods; in some cases the oostegites are reduced and the gills provide a brood chamber. A few euphausiaceans and one species of dendrobranchiate shrimp brood embryos on their posterior thoracopods. Most other decapods, including the carideans, stenopodideans, thalassinideans, astacids, palinurans, anomurans, and brachyurans, brood embryos on their pleopods.

Crustacean embryos proceed through some form of spiral cleavage, often highly modified by the amount of yolk present, to form a nauplius larva, a triangular shaped organism (larger in the anterior) with a simple median eye (ocellus) and three pairs of biramous appendages (the future antennules, antennae, and mandibles). The embryo may either emerge as a plankton-feeding nauplius when yolk reserves are depleted, or the nauplius may be laden with yolk, reducing the need to feed until a later stage. Larvae hatching as nauplii require long developmental times and are observed in many of the "less advanced" groups (some anostracans, some notostracans, some cladocerans, conchostracans [yolkey], some ostracods, mystacocarids [develop in interstitial sand], nonparasitic copepods [nonfeeding], some ascothoracidan cirripeds, some acrothoracican cirripeds, some thoracican cirripeds, rhizocephalan cirripeds [nonfeeding], euphausiaceans, and dendrobranchiate shrimps). Evolution appears to have progressed toward hatching at a more advanced stage, either as a metanauplius (more segments but still only three appendages), a protozoea (where a metanaupliar stage undergoes a metamorphic molt within the egg membrane to a larval stage with sessile compound eyes, maxillules, and maxillae) or a zoea (a more advanced larval form with stalked eyes and three pairs of maxillipeds; these swimming larvae eventually hatch into a postlarval form resembling an immature adult). The cephalocarids, fossil lipostracans, some anostracans, some notostracans, some cladocerans, some ostracods, some ascothoracidan cirripeds, and some acrothoracican cirripeds hatch as metanauplii. The reduction or loss of the free larval stage is accomplished by retaining the naupliar, metanaupliar, or protozoeal stages within the egg membrane during a longer brooding period before hatching or, in some cases, by shortening the total development time by omitting the morphological development of some or all of these larval stages entirely.

Crustaceans have three types of development, including "direct," "anamorphic," and "metamorphic" development; the last two forms sometimes are called less and more extreme types of "mixed" development. In direct development, embryos hatch as juveniles that resemble miniature adults, either because larval stages are passed inside brooded eggs before hatching, because they have been suppressed or lost, or possibly because they never were present. Direct development probably occurs in remipedes and probably was present in the Cambrian canadaspid malacostracans, either because larvae were suppressed or never present. Also, most cladocerans, most ostracods, most branchiurans, leptostracans, anaspidaceans and bathynellaceans, and all of the peracarids show direct development. The young hatch as an adult-like juvenile in some freshwater and terrestrial astacids and brachyurans and in some taxa of very small body size, including some stomatopods and some of the very small caridean shrimps. Most of the instances of abbreviated larval phases probably are explained by passage of larval stages within brooded eggs, but some cladocerans, some ostracods, branchiurans, and the anaspidacean and bathynellacean syncarids do not brood. Either larval stages were suppressed or never were present in these groups.

In anamorphic development, embryos hatch as some form of naupliar, metanaupliar, or later-stage larva, and the adult form is gradually achieved by a progression of morphological changes as the larva acquires more segments and appendages. Anamorphic development is found in many cephalocarids and allied fossil lipostracans, anostracans, notostracans, some cladocerans, conchostracans, some ostracods, mystacocarids, copepods, and one group of branchiurans. In the stomatopods, euphausiaceans and amphionidaceans, the morphology of the larvae generally resembles the form of the adult (thus not requiring a major metamorphosis) so these groups could be considered to have anamorphic development; however, these three groups usually are considered to have metamorphic development because they have protozoeal and zoeal larval forms.

In metamorphic development, the embryo hatches as some form of naupliar, metanaupliar, or later stage larva and develops with a body form that is radically different from that of the adult; thus, a major morphological reorganization (similar to that of holometabolous insects such as caterpillars and butterflies) is required before the adult stage can be attained. All of the eucarids as well as the stomatopods usually are considered to have metamorphic development because they possess protozoeal and zoeal larval forms. Other groups that undergo a major metamorphosis between the larval and adult phases include the cirripeds, although they lack the typical protozoeal and zoeal stages. The cirripeds are unique in developing a morphologically distinct nonfeeding bivalved postlarval form called a cyprid before transforming into the adult phase. In thoracicans, the cyprid attaches to the substratum with its antennule and reorganizes its body into a sessile adult. In addition to a cyprid stage, boring acrothoracicans metamorphose into a postcyprid pupa stage in order to reorganize itself into the adult form, similar to holometabolous insects. Following four feeding naupliar stages and a nonfeeding cyprid stage that attaches to the prospective host, the parasitic rhizocephalans form a hypodermic-like "kentrogon" stage through which the body, completely transformed, injects itself into the host's body. The parasitic tantulocarids and copepods also undergo metamorphoses in order to attain the highly modified adult parasitic form. The naupliar larval forms are suppressed in parasitic copepods, which hatch as copepodites (small juveniles).

The number of eggs produced by crustaceans generally correlates with body size, with larger species within lineages usually producing vastly more eggs than smaller species. This tracks only within lineages, however, and, if multiple lineages are combined, there is no trend. Certain crustacean groups are characterized by "weedy" life history characteristics while others are long lived and slow to reproduce. Anostracans and notostracans usually live only a few weeks and produce many eggs, including resistant eggs that are capable of surviving long periods of unfavorable conditions, similar to the seed banks of desert plants. Some copepods also produce dormant eggs. Cladocerans pass through several generations in one summer season. Amphionidaceans may be short-lived as adults (the female does not feed), and males in several taxa (gnathiid isopods, dwarf males in rhizocephalan cirripeds) are short-lived and do not feed. On the other hand, mysidans reproduce infrequently and may require up to 7 years to mature. Some Arctic cumaceans have delayed maturity and live many years, reproducing little. Some stomatopods, astacid and palinuran decapods, and some brachyurans live many years.

The significance of the type of development and number of eggs produced in crustaceans lies in their relationship to dispersal. Species that produce more eggs and those that are released at early stages so that longer periods are spent in the plankton have the largest geographic ranges and are the most resistant to environmental change and extinction. The resistant eggs of anostracans and notostracans often are dispersed by wind or on the feet of birds. Nonparasitic copepods develop through six naupliar stages and five copepodite stages. Stomatopods pass several protozoeal stages (advanced naupliar or metamaupliar stages) in the egg while brooded by the female. The protozoeal stage, termed an "antizoea" or "pseudozoea," is followed by four to eight or nine zoeal stages, called "erichthus" or "alima" larvae, in small reef and larger level bottom lineages, respectively. These modified zoeae are pelagic for periods that range from 3 to 4 weeks to 9 months in small reef versus larger level bottom species, respectively. A minority of small species of stomatopods produce only a few, unusually large eggs relative to their body size and undergo their entire larval development within the burrow of the female. Stenopodid shrimps, palinuran lobsters, and some anomurans are characterized by many larval stages and long pelagic phases. Most small species, most species that inhabit fresh water and terrestrial habitats, many species that inhabit high latitude or deep cold environments, and species with anamorphic or metamorphic development that hatch at advanced stages are likely to produce fewer, relatively larger eggs with shorter pelagic phases. These species, and those with direct development, are most likely to have extremely restricted, endemic distributions.

4. The Geographic Distributions of Crustaceans

Many of the geographic distributions for each of the major groups of crustaceans strongly reflect ancient conditions where the group originated, indicating that dispersal (usually by long lived pelagic larvae) has not dominated the historical ecology of many groups of crustaceans. The importance of intermediate dispersal in speciation, however, where larvae travel far enough to establish new populations but not sufficiently frequently to swamp arising genetic variants, is well illustrated in the thoracican barnacles. Also, the long lived pelagic phases of some of the large level bottom stomatopods have been shown to dampen diversification within and among regions, while those lineages with short lived larvae and significantly more restricted geographic ranges rapidly speciate and also suffer higher rates of apparent extinction (Reaka 1980; Reaka and Manning, 1981, 1987). The distributions of fossil taxa are not extensively considered below because, except for the worldwide distributions of the archaeostracan phyllocarid malacostracans, many extinct taxa are recorded from relatively few regions, possibly as a result of preservation bias.

To understand patterns of distribution in contemporary crustaceans, data for all of the major taxa were divided into nested categories where the entire group (subclass, order, or suborder) had a wide (>half the circumference of the earth) or narrow (≤half the earth's circumference) global distribution, those with many (>100) or few species (≤100), and those in which most of the individual species had wide or narrow geographic ranges, producing a table of eight columns. This procedure revealed the somewhat surprising result that entire lineages of crustaceans with narrow distributions are rare.

Restricted geographic distributions for entire lineages are found only in four taxa (the nectiopodan remipedes, anaspidacean syncarids, thermosbaenacean pancarids/peracarids, and spelaeogriphacean peracarids). Each of these groups also contains few species (10, 15, 11, and 2, respectively) that all have highly restricted individual geographic ranges (one or a few localities). These four lineages likely represent the remnants of once more broadly distributed groups that have experienced extinction outside of the marginal refugia they currently inhabit. There are no lineages with narrow distributions and few but broadly distributed species, and no lineages with narrow distributions and many species with either wide or narrow species ranges.

Among groups where the entire taxon has a broad distribution, only six groups have low species diversity and contain species with restricted geographic ranges (cephalocarids [10 species]; tantulocarids [5 species]; bathynellaceans [100 species]; mictaceans [3 species]; procaridid shrimps [3 species]; and stenopodidean shrimps [25 species]). Widely distributed lineages with few species but broad species ranges include another seven groups (notostracans [11 species], mystacocarids [10 species], acrothoracicans [most of the 50 species], leptostracans [most of the 13 species], lophogastridans [most of the 40 species], euphausiaceans [more than half of the 90 species], and the single, though variable, amphionidacean species).

Groups in which the entire taxon is widely distributed on a global scale and that contain many species with broad species ranges are not numerous. Only most dendrobranchiate shrimps and most palinurans fit clearly within this category, although whether the majority of copepods, cumaceans, anomurans, and brachyurans have wide or narrow individual species ranges is not well documented.

By far the most common distribution pattern observed in crustaceans includes lineages that are speciose (>100 species), broadly distributed on a global scale, and contain species with relatively restricted geographic ranges (14 groups, 18 if copepods, cumaceans, anomurans, and brachyurans, whose status is not certain, are included). These lineages include most of the ostracods (>8000 species), anostracans (>180 species), probably most of the cladocerans (>450 species), most of the conchostracans, (>180 species), most of the branchiurans (150 species), most of the thoracicans (>800 species), most of the stomatopods (>400 species), probably most of the tanaids (>850 species), most of the isopods (>4000 species), most of the amphipods (>6000 species), probably most of the carideans (2500 species), and probably most of the thalassinideans and astacideans.

In summary, the major lineages of crustaceans are most commonly broadly distributed on a global scale (29 of the 33 taxa documented). However, the dominant pattern for species is to have relatively restricted geographic ranges. In at least 24 of the 33 groups documented, the majority of species have relatively small geographic ranges, in some cases restricted to a single locality. Lineages containing many species and those with few species are about equally represented (16 and 17, respectively).

D. Vulnerability to Extinction

1. Extinction in the Fossil Record

Extinction in the fossil record tells us a considerable amount about how prone modern crustaceans are to extinction. Studies show that the breadth of geographic distribution of a species or lineage, which is intimately intertwined with their larval dispersal potential, is the single most important predictor of extinction. Factors that favor extinction or survival, however, depend on whether one is considering organisms during normal levels of background extinction or during cataclysmic mass extinctions (Jablonski, 1991). During background conditions, lineages with long-lived planktonic larvae persist longer in the fossil record (i.e., have lower extinction rates) than those with abbreviated development, individual species with broad geographic ranges suffer lower extinction than those with narrow endemic distributions, and lineages characterized by high species richness survive longer than those with few species. Species-rich lineages composed of broadly distributed species have especially high survival. During mass extinction episodes, however, all of these traits are ineffectual, and the broad geographic deployment of an entire lineage, regardless of the ranges of the individual species or the number of component species, is the primary characteristic that increases the likelihood that the lineage will persist into the next geological era (Jablonski, 1986, 1991).

Jablonski (1991) designated nine major mass extinctions (Late Ordovician, 61% and 85% of genera and species lost; Late Devonian, 55% and 82% of genera and species lost; Late Permian, 84% and 96% of genera and species lost; Late Triassic, 47% and 76% of genera and species lost; Middle Jurassic, 26% and 53% of genera and species lost; End Jurassic, 21% and 45% of genera and species lost; Middle Cretaceous 26% and 53% of genera and species lost; End Cretaceous 47% and 76% of genera and species lost; Late Eocene, 15% and 35% of genera and species lost). Of 15 crustacean lineages that became extinct, 9 of them did not persist beyond one of these biotic crises. Although this does not prove that the mass extinctions were directly responsible for their demise, the failure of lineages to survive into the next period, especially in taxa that had persisted through several previous geological periods, suggests that the conditions associated with the mass extinction may have contributed to their disappearance. In many cases, representation in the fossil record is too spotty to determine whether the lineage and its component species had broad geographic ranges, but some useful information can be gleaned from what we know about fossil crustaceans. The lipostracans (allied to the cephalocarids) and the archaeostomatopod hoplocarids are known only from the Devonian, and the kazacharthrans (allied to the notostracans) are recorded only from the Jurassic. Mass extinctions closed both of these periods, and it appears that these lineages were not widely distributed, being known only from Laurentia (North America and Europe), Scotland, and Kazachstan, respectively. Also, only a few species are known in each lineage. Known only from equatorial Laurentia in the Cambrian and Ordovician, the hymenostracan phyllocarid malacostracans did not survive beyond the Late Ordovician mass extinction. The extraordinarily successful archaeostracan phyllocarid malacostracans persisted from the beginning until the end of the Paleozoic, making it through both the Late Ordovician and

Late Devonian crises but failing to survive the greatest of mass extinctions in the Late Permian. This lineage had a worldwide distribution and included many taxa that occupied diverse habitats; these factors may have facilitated their persistence through several major crises but were not sufficient to allow survival through the End of Paleozoic debacle. Other primitive malacostracans, including the waterstonellideans, eocaridaceans, paleocaridacean syncarids, and pygocephalomorph mysidaceans, arose in the Carboniferous but also perished during the End of Paleozoic extinction. On the other hand, however, one order of the primitive remipedes (the enantiapodans), the canadaspid and hoplostracan phyllocarid malacostracans, the aeschronectid hoplocarids, and the belotelsonid malacostracans, all known only from the Carboniferous, and the palaeostomatopod hoplocarids, which originated in the Devonian and persisted until the Carboniferous, all perished in the absence of a major biotic crisis.

Tropical, especially reef, biotas have throughout history suffered disproportionately greater losses to mass extinctions than those at higher latitudes (Jablonski, 1991). Perusal of the habitats and distributions of the extinct fossil crustaceans suggests that most of those lost occurred in regions that were tropical at the time. The tropics also are the largest source of evolutionary novelties (due to speciation and development of major new adaptive types). This is possibly because the recurring extinction events open new evolutionary opportunities there, because high diversity in the tropics makes origination of new types more probablistic simply because of the large numbers of lineages available for genetic experiments, or because something about the ecology of tropical environments causes greater divergence per speciation event (Jablonski, 1993).

Low density and small total population size also has been shown to render species susceptible to extinction in some taxa (Raup, 1986), and this characteristic has been of overriding concern to many conservation biologists in recent years. Evidence on the importance of population density as a predictor of extinction is equivocal in crustaceans. On one hand, the archaeostracan phyllocarid malacacostracans were one of the most abundant benthic organisms, and they persisted from the Cambrian until the end of the Paleozoic, supporting the thesis that population density is important. On the other hand, however, the archaeostomatopod hoplocarids were abundant during the Devonian but did not survive into the next period. Aeschronectid hoplocarids dominated their benthic communities and belotelsonid malacostracans were abundant during the Carboniferous, but neither persisted beyond that period. Waterstonellidean malacostracans formed abundant aggregations, palaeocaridacean syncarids were very numerous, and pygocephalomorph mysidaceans were among the most important components of their communities during the Late Paleozoic (Carboniferous to Permian); yet all succumbed to extinction at the end of the Permian.

Similarly, when a species occupies a unique or particularly restricted type of habitat, environmental changes are more likely to entirely eliminate these localized populations, and individuals may be unable to find or migrate to a similar specialized environment. This feature of organism-environmental interaction also has been a dominant concern in conservation biology. The evidence from fossil crustaceans on this topic, however, is not clear. An example of a lineage with a very specialized habitat that became extinct is the lipostracans, allied to the cephalocarids, that lived in Devonian hot springs. However, other crustaceans that became extinct were not obviously specialized for unique habitat types and inhabited apparently generalized benthic coastal environments. Enantiopodan remipedes; canadaspid and hoplostracan phyllocarid malacostracans; belotelsonid and eocaridacean malacostracans; archaeostomatopod, aeschronectid, and palaeostomatopod hoplocarids; palaeocaridacean syncarids; and pygocephalomorph mysidaceans all inhabited shallow benthic coastal environments. Of these, at least the abundant aeschronectids, pygocephalomorphs, and palaeocaridaceans also co-occurred in brackish swamps and deltas, and at least the palaeocaridaceans extended into fresh water. The latter four lineages possibly could be considered to exhibit some habitat specialization (fresh water), but most of the lineages lost to extinction were not habitat specialists.

Broadly compared across phyla, the fossil record shows that species high on the food pyramid (predators or parasites) have sometimes been at risk during environmental perturbations, particularly if they also had achieved large body size and had developed specialized morphology and high levels of activity (e.g., for predation). It is thought that their high metabolic requirements and need for large amounts of food imperils their survival during difficult periods compared to smaller species that usually are lower in the food web. A related concept, known from the 1800s as Cope's Rule, posits that most lineages originate as small generalized forms, evolve toward larger and more specialized morphologies, then become extinct (Stanley, 1973). Body sizes of extinct crustaceans are not synthesized in the literature in most cases, but the few available cast doubt on whether or not crustacean lineages of large body size are particularly succeptible to extinction. Enantiopodan remipedes probably were relatively small, kazacharthrans were small to moderately sized (<30 mm), and palaeocaridaceans and phygocephalomorphs probably were small to moderately sized for crustaceans. Yet all became extinct. In many crustaceans (Reaka and Manning, 1981, 1987; Reaka, 1980) and among invertebrates in general (Menge 1975; Strathmann and Strathmann, 1982), larger species produce many more small offspring that spend considerably longer periods in the plankton and have wider geographic ranges than smaller species. Larval dispersal and wide geographic ranges tend to slow down both extinction and speciation (Jablonski 1986; Jablonski and Lutz, 1983), suggesting that, in marine invertebrates, dispersal at large body sizes usually overpowers susceptibility to extinction. However, not all large-sized species produce large numbers of widely dispersing offspring (especially in brackish, freshwater, and terrestrial environments or in lineages where abrreviated development is inherited from past legacies), and available food may be sufficiently restricted to prevent reproduction (thus dispersal) during particularly stressful environmental conditions. Thus, although Cope's Rule may apply more universally among fossil vertebrates (which were Cope's original source of data for the pattern of greater extinction in species of large body size) than among invertebrates, large body size still may render lineages susceptible to extinction among invertebrates in some situations.

2. The Prospects of Extinction in Modern Crustaceans: Which Taxa Are Most Vulnerable, and Which Ones Represent the Most Urgent Conservation Targets?

Table II summarizes the major factors that are likely to affect the viability of contemporary crustacean groups during natural or human-induced environmental changes. Many scientists caution that drastic environmental shifts could occur as human populations increase exponentially (perhaps doubling in the next 100 years) and affect the atmosphere, soil, and coastal waters (Wilson, 1989). If such change impacts the biotic world in a manner similar to the mass extinctions of the past, the primary factor that would facilitate survival of the crustacean groups analyzed in this study would be wide distribution of entire groups. Fortunately, most (35 of 39) groups of crustaceans are distributed nearly worldwide, suggesting that at least some of these groups probably would survive. The remipedes, anaspidaceans, thermosbaenaceans, and spelaeogriphaceans, however, would be at high risk.

If the deteriorating environment resembled background conditions rather than a mass extinction event, however, the number of lineages within the group, the size of the geographic ranges of species, whether or not most of the group inhabited tropical (especially reef) environments, and possibly whether the species were characterized by low population densities, occupied specialized or highly restricted types of habitat, were positioned high in the food web of their communities, or were large and active could influence the probability that they would succumb to extinction over the long run.

Of the 34 groups for which data were available on species numbers, 25 groups (74%) were considered to be at risk as a result of relatively few species within the lineage (14 at high risk, 11 at moderate risk). The remipedes, cephalocarids, notostracans, mystacocarids, leptostracans, tantulocarids, anaspidaceans, thermosbaenaceans, and spelaeogriphaceans all have fever than 15 species and would be at very high risk. Even more notable, the mictacean and procaridid lineages are represented by only 3 species each, and only 1 species is known in the entire Order Amphionidacea. These phylogenetically unique groups could easily be lost.

Of 39 groups, 27 (70%) are composed of species with mostly restricted geographic ranges. Ten of these groups contain species with especially restricted distributions, considered at high risk, and 17 contain some species with restricted ranges that are judged to be at moderate risk. Of the 39 groups, 14 (36%) are considered to be at risk because they are primarily tropical or reef dwelling. Twenty-seven (74%) of the 39 groups are characterized by small population sizes. Eight of these have especially low densities or total populations and are considered to be at high risk, and 19 appear to be at moderate risk. Of the 39 groups, 29 (74%) occupy unique or highly specialized types of habitats; 11 and 18 of these are judged to be at high or moderate risk, respectively, because of their degree of habitat specialization. Although many crustaceans are predators or parasites, relatively few are at the apex of the food web (this position often being usurped by fishes), and relatively few are species specific in their choice of prey or hosts. Of the 39 groups, 23 (59%) are judged to be at moderate risk due to a relatively high position in the food chain. Also, relatively few entire groups of crustaceans are sufficiently large in body size that their persistence is likely to be imperiled. Large body size is unlikely to be associated with extensive extinction for the following reasons. Wherever the distributions of

TABLE II

Summary of Extinction Vulnerability and Conservation Value of Different Groups of Contemporary Crustaceans

Taxon	Few living species (++<100, +<500)	Entire group restricted geographic range (<half circumference Earth)	Most species restricted geographic range	Group primarily tropical	Small population size	Unique or restricted habitat type	High trophic level	Large body size (or long lived)	Old (+probably since mid-Paleozoic, ++probably since beginning Paleozoic) or phylogenetically unique lineage	Total score (sum of +,± counted as 0.5)
Remipedia	++(10)	++	++	+	++	++(brackish caves)	+		++	15
Cephalocarida	++(10)		++		++				+	7
Anostraca	+(≥180)		+			+(ephemeral pools)			+	4
Notostraca	++(11)					+(ephemeral pools)			+	4
Cladocera	+(≥450)		+?							1.5
Conchostraca	+(≥180)		?						+	2
Ostracoda	(about 8000)		++	+		+(patchy)	±(some)		++	6.5
Mystacocarida	++(10)				±(patchy)	++(interstitial)	±(some)			5
Copepoda	(9000)		?	+	±(some)	+(some parasites)	±(some)	±(some parasites, relative)		3.5
Branchiura	+(150)		+	+		+(ectoparasites)	+			5
Tantulocarida	++(5)		++		++	++(parasites)	+	±(relative)		9.5
Ascothoracida	+(60)		?	+	+	+(parasites)	+		++	7
Acrothoracica	+(50)			+	+	+(borers)			+	5
Thoracica	(≥800)		+		±(some)	+(intertidal, commensal)			++	4.5
Rhizocephala	+(230)		?		±(local)	++(endoparasites)	+	+(relative)		5.5
Leptostraca	++(14)		+		+				+	5
Stomatopoda	+(400)		+	+	±(some)	±(some)	+(active)	±(some)	+	6.5
Anaspidacea	++(≥15)	++	++		+	++(fresh water refugia)			±(allied to primitive groups)	9.5
Bathynellacea	+(100)		++		++?	++(ground water)			±(allied to primitive groups)	7.5
Mysida	(>745)		+		±(some)	+(some intertidal, interstitial, fresh water, caves)	±(some)	±(some long lived, low reproduction)	±(allied to primitive groups)	4
Lophogastrida	+(40)		?		+?		±(some)		+	3.5
Thermosbaenacea	++(11)	+	++		+	++(fresh and salt ponds, caves, hot springs, ground water)			+	9
Tanaidacea	(850)		+?				±(some)		+	2.5
Spelaeogriphacea	++(3)	+	++		++	++(fresh water caves)			+	10
Cumacea	(≥1000)		?		±(patchy)	±(patchy)	±(some)	±(some long lived)	+	3
Mictacea	++(3)		++		++	++(caves, deep sea)			+	9
Isopoda	(>4000)		+		±(some)	+(many, including terrestrial, parasites)	±(some)		+	4
Amphipoda	(>6000)		+			±(some, including terrestrial, parasites)	±(some)			2
Euphausiacea	++(90)		±(some)				+			3.5
Amphionidacea	++(1)				++				++phylogenetically unique, 1 genus and species	6
Dendrobranchiata	+(450)			+			+(active)	+		4
Caridea	(>2000)		+	+	±(some)	±(some)	+(active)	±(some)		5.5
Procarididae	++(3)		++		+	++(brackish caves)			+	8
Stenopodidea	++(25)		+	+	++	++(commensal)	+(active)			9
Thalassinidea			+	+	±(some)	±(some, cryptic reef)	+(some, active)	+(many)		10
Astacida			+			+	+(active)	+	+	5
Palinura			±(some)	+	+		+(active)	+	+	5.5
Anomura				+		+(many)	±(some, active)			2.5
Brachyura			±(some)	+	±(some)	+(many)	+(many, active)	+	+?	6

body sizes among species of a group have been compared, there always are far more species of small than large body size (May 1988; Reaka-Kudla, 1991, 1997), and this almost certainly is true throughout the Crustacea. Also, in many crustacean groups, species of large body size produce long-lived pelagic larvae that confer protection against background extinction. The present analysis suggests that 12 of 39 groups (31%) are at possible risk due to their relatively large body sizes (if other life history characteristics do not outweigh the effect of body size).

One of the factors that should be included in prioritizing conservation efforts is the uniqueness of the lineage. Some crustacean groups represent a heritage that stems from half a billion years ago (the beginning of the Paleozoic). Among groups still alive today, the remipedes, ostracods, ascothoracidan barnacles, and thoracican barnacles are known from Cambrian fossils or their morphology suggests that they are derived from the most primitive lineages of crustaceans. Sixteen other groups are known from the Mid-Paleozoic (300–400 million years ago). These ancient groups are more likely than more recently derived groups to have totally novel characteristics, even in their biochemical and genetic constitution. As we enter a millenium in which biotechnology has only begun to offer human benefits through genetic manipulations, we can ill afford to lose these lineages. Some of these ancient groups are particularly vulnerable due to few constituent species, restricted geographic ranges, small populations, specialized habitat requirements, or other factors (especially the remipedes, cephalocarids, notostracans, ascothoracidan and acrothoracican barnacles, leptostracans, anaspidaceans, thermosbaenaceans, spelaeogriphaceans, mictaceans, procaridids, and especially the amphionidaceans). These groups merit special conservation attention.

Table II provides a tentative overall summation score for risk factors and intrinsic conservation value in each group. The overall mean risk/value score for the 39 groups was 5.9. Incorporating all of the different risk and value factors, eight groups (the remipedes, tantulocarids, anaspidaceans, spelaeogriphaceans, mictaceans, procaridids, stenopodideans, and thalassinideans) merit the closest conservation scrutiny overall (score of 9 or above). Of these, the remipedes are considered to be the highest priority. This represents only a beginning, however, at assessing which lineages are in most need of protection in order to ensure their survival in the next centuries. Identifying which lineages and where they are most at risk (as well as the most effective methods for conserving them) will be one of the most important tasks for at least the next decade.

See Also the Following Articles

AQUACULTURE • ENDANGERED MARINE INVERTEBRATES • INVERTEBRATES, MARINE, OVERVIEW • MARINE ECOSYSTEMS • MARINE ECOSYSTEMS, HUMAN IMPACT ON • MOLLUSCS

Bibliography

Abele, L. G. (Ed.) (1982a). *The Biology of Crustacea.*. Vol. 1: *Systematics, the Fossil Record, and Biogeography.* Academic Press, New York.

Abele, L. G. (Ed.) (1982b). *The Biology of Crustacea.*. Vol. 2: *Embryology, Morphology, and Genetics. Academic Press, New York.*

Barnes, R. D. (1980). *Invertebrate Zoology*, 4th ed. Saunders, Philadelphia.

Barnes, R. S. K. (Ed.) (1998). *The Diversity of Living Organisms.* Blackwell Sci. Ltd., London.

Bowman, T. E., and Abele, L. G. (1982). Classification of the Recent Crustacea. In *The Biology of Crustacea.* 1. *Systematics, the Fossil Record, and Biogeography* (L. G. Abele, Ed.), 1–27. Academic Press, New York.

Brusca, R. C., and Brusca, G. J. (1990). *Invertebrates.* Sinauer Assoc., Sunderland, MA.

Dominguez, J. H., and Reaka, M. L. (1988). Temporal activity patterns in reef-dwelling stomatopods: A test of alternative hypotheses. *J. Exp. Mar. Biol. Ecol.* **117**, 47–69.

Feldmann, R. M., and Manning, R. B. (1992). Crisis in systematic biology in the "Age of Biodiversity". *J. Paleontol.* **66**, 157–158.

Gaston, K. J., and May, R. M. (1992). Taxonomy of taxonomists. *Nature* **356**, 281–282.

Gore, R. H., and K. L. Heck. 1986. *Crustacean Biogeography.* Balkema Press, Rotterdam.

Grassle, J. F., and Maciolek, N. J. (1992). Deep-sea species richness: Regional and local diversity estimates from quantitative bottom samples. *Amer. Nat.* **139**, 313–341.

Jablonski, D. (1986). Background and mass extinctions: The alternation of macroevolutionary regimes. *Science* **231**, 129–133.

Jablonski, D. (1991). Extinctions: A paleontological perspective. *Science* **253**, 754–757.

Jablonski, D. (1993). The tropics as a source of evolutionary novelty through geological time. *Nature* **364**, 142–144.

Jablonski, D., and Lutz, R. A. (1983). Larval ecology of marine benthic invertebrates: Paleobiological implications. *Biol. Rev.* **58**, 21–89.

Knowlton, N. (1993). Sibling species in the sea. *Ann. Rev. Ecol. Syst.* **24**, 189–216.

Lovejoy, T. E. (1980). Changes in biological diversity. In *The Global 2000 Report to the President* 2 (Tech. Rept.), (C. O. Barney, Eds.), pp. 327–332. Harmondsworth, Penguin Books.

May, R. M. (1988). How many species are there on Earth? *Science* **241**, 1441–1449.

May, R. M. (1994). Biological diversity: Differences between the land and sea. *Phil. Trans. Roy. Soc. Lond.* **B 343**, 105–111.

Menge, B. A. (1975). Brood or broadcast? The adaptive significance of different reproductive strategies in two intertidal sea stars, *Leptasterias hexactis* and *Pistaster ochraceus*. *Mar. Biol.* **31**, 87–100.

Norse, E. A., and McManus, R. E. (1980). Ecology and living resources: Biological diversity. In *Environmental Quality 1980: The Eleventh Annual Report of the Council on Environmental Quality.* pp. 31–80. Council on Environmental Quality, Washington, D.C.

Pearse, V., Pearse, J., Buschbaum, M., and Buschbaum, R. (1987). *Living Invertebrates.* Blackwell Sci. Pub., Palo Alto, CA.

Poore, G. C. B., and Wilson, G. D. F. (1993). Marine species richness. *Nature* **361**, 579.

Raup, D. M. (1986). Biological extinction in Earth history. *Science* **231**, 1528–1533.

Ray, G. C. (1991). Coastal zone biodiversity patterns. *Bioscience* **41**, 490–498.

Reaka, M. L. (1980). Geographic range, life history patterns, and body size in a guild of coral-dwelling mantis shrimps. *Evolution* **34**, 1019–1030.

Reaka, M. L., and Manning, R. B. (1981). The behavior of stomatopod Crustacea, and its relationship to rates of evolution. *J. Crust. Biol.* **1**, 309–327.

Reaka, M. L., and Manning, R. B. (1987). The significance of body size, dispersal potential, and habitat for rates of morphological evolution in stomatopod Crustacea. *Smithsonian Contr. Zool.* **448**, 1–46.

Reaka-Kudla, M. L. (1991). Processes regulating biodiversity in coral reef communities on ecological vs. evolutionary time scales. In *The Unity of Evolutionary Biology* 1 (E. C. Dudley, Ed.), pp. 61–70. Dioscorides Press, Portland, OR.

Reaka-Kudla, M. L. (1997). The global biodiversity of coral reefs: A comparison with rain forests. In *Biodiversity II: Understanding and Protecting Our Natural Resources* (M. L. Reaka-Kudla, D. E. Wilson, and E. O. Wilson, Eds.), pp. 83–108. Joseph Henry/National Academy Press, Washington, D.C.

Reaka-Kudla, M. L., Wilson, D. E., and Wilson, E. O. (1997). Santa Rosalia, the turning of the century, and a new age of exploration. In *Biodiversity II: Understanding and Protecting Our Natural Resources* (M. L. Reaka-Kudla, D. E. Wilson, and E. O. Wilson, Eds.), pp. 507–524. Joseph Henry/National Academy Press, Washington, D.C.

Schram, F. R. (Ed.), (1983). *Crustacean Phylogeny*. Crustacean Issues I. Balkema Press, Rotterdam.

Schram, F. R. (1986). *Crustacea*. Oxford University Press, Oxford.

Stanley, S. M. (1973). An explanation for Cope's Rule. *Evolution* **27**, 1–26.

Strathmann, R. R., and Strathmann, M. F. (1982). The relationship between adult size and brooding in marine invertebrates. *Amer. Nat.* **119**, 91–101.

Systematics Agenda 2000. (1994). *Systematics Agenda 2000: Charting the Biosphere*. Tech. Rept. Consortium of Amer. Soc. Plant Taxonomists, Soc. Syst. Biol., Willi Hennig Soc., in cooperation with Assoc. Syst. Collections, New York.

Wilson, E. O. (1985). Time to revive systematics. *Science* **230**, 1227.

Wilson, E. O. (1989). Threats to biodiversity. *Sci. Amer.* (September), 108–116.

Wilson, E. O. (1992). *The Diversity of Life*. Belknapp Press, Cambridge, MA.

Wilson, E. O., and Peters, F. M. (Eds.) (1988). *BioDiversity*. National Academy Press, Washington, D.C.