Essential Fungal Genetics

David Moore
LilyAnn Novak Frazer

SPRINGER

Essential Fungal Genetics

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

David Moore LilyAnn Novak Frazer

Essential Fungal Genetics

With 59 Illustrations

Springer

David Moore
School of Biological Sciences
The University of Manchester
Manchester M13 9PT
UK
david.moore@man.ac.uk
www.oldkingdom.org

LilyAnn Novak Frazer
Response Environmental Technologies Limited
Abbeyfield House
Blyth Road
South Yorkshire S66 8HX
UK
Lily.Nfrazer@response-uk.com

Library of Congress Cataloging-in-Publication Data
Moore, D. (David), 1942–
Essential fungal genetics/David Moore, LilyAnn Novak Frazer.
p.;cm.
Includes bibliographical references and index.
ISBN 0-387-95367-1 (h/c: alk. paper)
1. Fungi—Genetics. I. Frazer, LilyAnn Novak, 1965– II. Title.
[DNLM: 1. Fungi—genetics. 2. Genome, Fungal. QK 602 M821e 2002]
QK602 .M66 2002
579.5′135—dc21 2001054910

Printed on acid-free paper.

Production coordinated by Chernow Editorial Services, Inc., and managed by Francine McNeill; manufacturing supervised by Jerome Basma.
Typeset by SNP Best-set Typesetter Ltd., Hong Kong.
Printed and bound by Maple-Vail Book Manufacturing Group, York, PA.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95367-1 SPIN 10853243

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Preface

Our intention in writing this book is to provide a text that offers a succinct but comprehensive account of the genetics of Kingdom Fungi. We do not pretend that this is a monographic textbook, nor do we intend it to be a laboratory manual. Rather, we hope that it will be a user-friendly guide that can be used as a supplement to general genetics texts.

Most genetics textbooks deal adequately with plant and animal genetics. Even at an elementary level, a description of Mendel's experiments with peas is rapidly followed by the description of Morgan's experiments with the fruit fly. Thus, right from the start plant and animal work is well integrated in the learner's mind.

This treatment unfortunately leaves the student ignorant of an entire kingdom of organisms. There are three major kingdoms of eukaryotes on Earth: plants, animals, and fungi. For the most part, fungi feature regularly in only two areas of traditional genetics teaching. On the one hand, we can find the ascus segregations that contributed so much to developing an understanding of the mechanism of recombination in the 1960s. On the other hand is the contribution work on yeast (as a model eukaryote) is currently making to understanding "cell cycle control" and its genetic regulation. As a result, most introductory genetics texts will leave the reader/student with the impression that fungi are of use when peculiarities of their structure or lifestyle suit them to particular experimental approaches, but are not worth mentioning otherwise!

We cannot redress the balance totally, but we hope we have produced a book that displays the genetics of fungi in a way that will be attractive and challenging, comprehensive yet succinct. This book will provide the tools for integrating fungal genetics into current teaching by complementing the major textbooks used in courses on general genetics, general organismal biology, general microbiology, and general mycology.

Our aim is to preserve the concepts that characterize fungal genetics, revealing the mixture of facts, techniques, and experimental approaches that distinguish fungal genetics as a study in its own right. We believe we have incorporated the full range of genetic information so that all aspects

of eukaryote genetics appropriate to fungi are described here in a way that is easily understood and memorable.

Without attempting to be a laboratory manual, this book has an instructional "how to do it" tone and will therefore have a practical value for anyone who has a mind to start genetic analysis of any fungus. Instead of dwelling on technical details like laboratory recipes, this book deals with the details of the ideas and concepts underlying basic approaches to fungal genetics so that these are on record and easily accessible rather than being lost in history, or buried by the great mountains of molecular methodologies.

We have tried to maintain a readable style of presentation, and to that end, there are no reference citations in the text; however, we provide a range of reference materials at the end of each chapter, and the number of references increases in later chapters as we touch on wider topics. We have favored review articles and Websites for these reference lists, so that the interested reader can very quickly penetrate deeper into the literature. More recent publications are also emphasized, but we also recognize the value of the history of the science and give references to some historical publications. For some of us (DM included), "history" is unfortunately deemed to start in about the 1980s. Doesn't time fly!

Another unusual aspect is that the examples we have chosen to illustrate various features tend, whenever possible, to favor basidiomycetes and the less-often referenced fungi. We have done this because it is relatively easy to find information about the main "models," like *Neurospora* and *Aspergillus*, so we prefer to describe other species. Of course, we can't escape "yeast," the ultimate model eukaryote, and the first eukaryote to have its genome completely sequenced. We try to emphasize, however, that no matter how much we learn about *Saccharomyces cerevisiae*, it is not the end of the story. No matter how important it is as a "model," this yeast is not adequately representative of filamentous fungi, nor does it represent other yeasts. When you know everything there is to know about *S. cerevisiae*, you will still be ignorant about fungal genetics. So read on, and get an impression of the full story!

David Moore
LilyAnn Novak Frazer
Manchester, United Kingdom
August 2001

Contents

This page intentionally left blank

CHAPTER 1

Why Study the Genomes of Fungi?

Revision Concepts

- Animals, fungi, and plants form three distinct kingdoms.
- Eukaryotes and eubacteria last shared a common ancestor 2×10^9 years ago.
- The three eukaryote kingdoms diverged 1×10^9 years ago, with the plant kingdom diverging from the other two eukaryote kingdoms first.
- The manner in which fungi gain nutrients distinguishes them from the other two eukaryote kingdoms.
- The three eukaryote kingdoms also manage their development, shape, and form differently.
- Development in fungi is constrained by the apical growth of the hypha and the formation of septa between cells at right angles to the axis of growth.
- Fungi are not well-represented by the fossil record.
- Evolutionary trees are mainly based on molecular structural evidence.
- True fungi form a monophyletic clade, which means that they all share a common ancestor.
- Initial evolution of land plants may have occurred as a result of cooperative associations between fungi and plants.
- Oldest terrestrial fossils of thread- and tubelike structures suggest the first terrestrial organisms were actually fungi.
- Phyla in the kingdom fungi include Chytridiomycota, Zygomycota, Ascomycota, and Basidiomycota.
- Chytridiomycota is the only fungal group that has flagella at any stage in its life cycle.
- Zygomycota have coenocytic hyphae, but no flagella.
- Ascomycota are the most numerous fungi, distinguished by the ascus (= meiocyte) and including the most well known: the yeast *Saccharomyces cerevisiae.*
- Basidiomycota comprise the "mushroom" fungi and are distinguished by the basidium (= meiocyte) and a long dikaryotic phase.

- Fungi contribute to nutrient recycling, form mutually beneficial symbioses with plants, and contribute to the food web.
- Commercial fungal products include alcohol, citric acid, and various drugs and medicines.
- Fungi, such as the yeast *S. cerevisiae*, are model organisms because of their small but representative eukaryotic genome.
- The genome is the entire DNA content of the cell.
- Prokaryotic and eukaryotic genomes differ in their character, but their machinery has a great deal in common.
- Obtaining the DNA sequence of a genome is the prelude to intensive analysis, involving the science of bioinformatics.

1.1 Origins

Fungi are eukaryotes, just like animals and plants. Although they have the same basic genetic structures as do animals and plants, there are some differences arising from the fungal lifestyle that need special explanation, which is one reason why they merit special study. There are other reasons that make fungi worthy of investigation in their own right, however, which we will explain in what follows. In summary, these reasons are: (1) the part played by fungi in nature make them crucially important to the maintenance of life on Earth; (2) they are enormously useful to us in industry right now, and they have an enormous potential for us to exploit in the future; and (3) as probably the oldest evolutionary line of eukaryotes, they provide us with easily studied model organisms.

For many years people thought that fungi were plants. Peculiar plants, admittedly, because it had to be assumed that they had lost photosynthesis during their evolution, but in the bipartisan view of the living world that held sway for so long, if they could not be called "animals" then they had to be "plants." It was realized in the mid-1960s that this was wrong and that animals, fungi, and plants form *three* distinct kingdoms of eukaryotes, bringing the total number of kingdoms defining the major groups of organisms up to five.

The most fundamental division of living organisms is into eukaryotes, the cells of which contain membrane-bound organelles (e.g., the nucleus and mitochondria), and the prokaryotes, whose cells lack extensive internal membrane-bound compartments. There are two very different groups of prokaryotes: eubacteria, which include most of the commonly encountered bacteria such as *Escherichia coli* and *Bacillus subtilis*, and the Archaea, which might be evolutionary relics of the earliest life forms on Earth and are mostly found in extreme environments like hot springs, brine pools, and anaerobic sediments. Eubacteria and Archaea are placed in Kingdom Prokarya; all the eukaryotes belong to the other four kingdoms. Kingdom Protoctista is a diverse collection of mostly single-celled microbes like cil-

iates and algae. The "higher organisms" that are able to organize their cells into multicellular tissues of particular function are placed in the remaining three kingdoms: Kingdom Plantae for all the green plants, Kingdom Animalia for the animals, and the fifth kingdom, Kingdom Fungi. Arranging organisms into groups and giving them scientific names (i.e., binomials, comprising a generic name and a specific name, like *Homo sapiens* or *Saccharomyces cerevisiae*) is called *systematics*, an agreed scheme of naming things. If the classification scheme is natural (rather than artificial), then the hierarchy of names will relate to the evolution of the organisms (i.e., their phylogeny).

When did it all start? The solar system formed about 4.5×10^9 years ago. There are microbial fossils in terrestrial rocks that are 3.5×10^9 years old. Life might have evolved even before that time, but calculations based on study of craters on the Moon suggest that the Earth–Moon system was subjected to gigantic asteroid impacts up to about 3.8×10^9 years ago. These massive impacts would have released enough energy to heat-sterilize the Earth's surface. If any life had evolved in those more distant times it would have been destroyed by the next impact. Once these cataclysmic impacts stopped and the Earth's surface stabilized sufficiently for life to evolve, the first bacterialike fossils would have been laid down. After this, there was a period of 1.5×10^9 years during which early bacteria continued to evolve before the higher organisms emerged. Even though eukaryotes and eubacteria last shared a common ancestor about 2×10^9 years ago, the three eukaryotic kingdoms diverged from one another about 1×10^9 years later. Studies of molecular evolution done in the late 1990s show that, surprisingly, animals and fungi are each other's closest relatives.

The sequence that emerges is that plants arose from the common eukaryotic ancestor 1×10^9 years ago, then a joint fungal–animal line continued for another 200 million years until that lineage diverged 800 million years ago. Recognizable fungi must have been around as long ago as that because from rocks only a few hundred million years younger, about 570 million years old, we find evidence in the form of fossil spores for all the major groups of fungi that exist today. It is quite clear that fungi were crucially important in the shaping of ancient ecosystems.

A major part of the original 1960s definition of the Kingdoms was their nutrition. Plants use the direct radiant energy from the sun to make their food; animals, from worms to killer whales, engulf their food. Fungi degrade food externally and absorb the nutrients that are released. One of the original articles on this topic emphasized the point: ". . . nutritive mode and way of life of the fungi differ from those of the plants . . . Fungi characteristically live embedded in a food source or medium, in many cases excreting enzymes for external digestion, but in all cases feeding by absorption of organic food from the medium. Their organization . . . is adapted to this mode of nutrition." Once this apparently simple basis for making the grand separation between Kingdoms has been used, numerous other differences

in structure and lifestyle fall into place. In particular, the three "higher" Kingdoms are very different from one another in ways that are crucial to determining shape and form. This is presumably because the evolutionary separation between the major Kingdoms must have occurred at a stage when the most highly evolved things were single cells. Thus, each Kingdom probably evolved the management of populations of cells independently of the others.

Cell migration (and everything that controls it) plays a central role in the embryological development of animals. Plant cells are encased in walls, and so they have little scope for movement. Changes in shape and form in plants are achieved by regulating the orientation and position of the wall that forms when a plant cell divides. Fungi also have strong cell walls, but their basic structural unit is a tubular, threadlike cell called a *hypha*. The hypha has two peculiarities that make fungal development totally different from that in plants: the hypha grows only at its tip and new walls form only at right angles to the growth axis. Fungal development, therefore, depends on control of the positioning and mode of growth of hyphal branches. This is a unique developmental system that poses unique evolutionary problems.

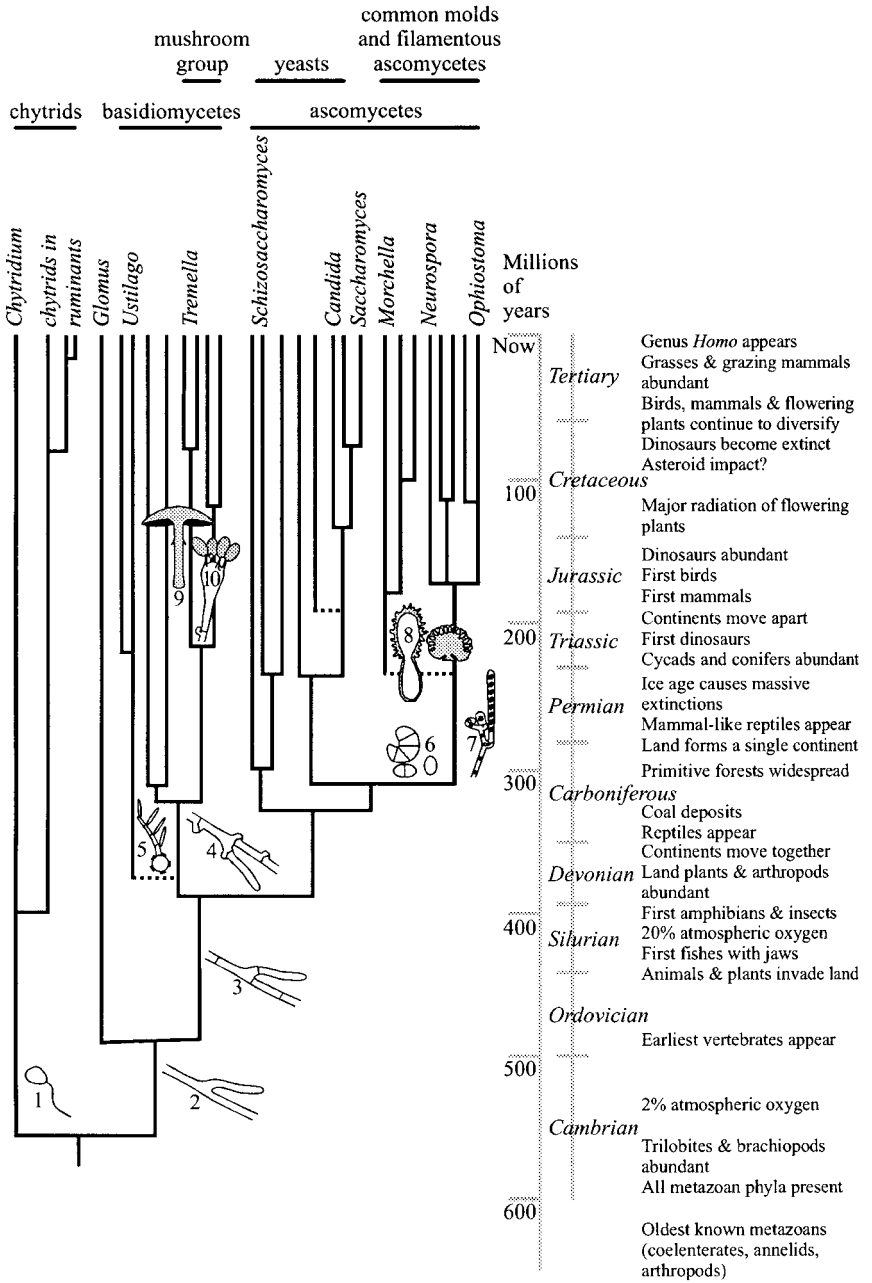
The evolution of fungi cannot be established from a good collection of fossils. There are some fossils, which we will mention later, but they are relatively few and scattered across evolutionary time. This has meant that we have to use evolutionary trees constructed from analysis of molecular (protein and nucleic acid) structures. This approach has been validated by comparisons in organisms (like many animal groups) for which a good fossil record does exist. Population genetics and molecular biology have been brought together with ever-increasing computational power into the field known as *molecular evolution*. DNA is the fundamental unit of molecular evolution, and rates and patterns of change in DNA and protein sequences can allow inference of the evolutionary history of organisms. The approach works well if you don't try to extract too much detail from it. Thus, we will restrict ourselves to the major messages that come out of this work in what follows.

The most important message is that true fungi, the Kingdom Fungi with its four constituent phyla, form what is known as a *monophyletic clade*. A *clade* is a group of organisms that share a common ancestor. *Monophyletic* refers to any group of organisms that are assumed to have originated from the same, common, ancestor. Thus, the molecular data indicate that true fungi all have a common origin. There are then some separate clades of other organisms that mycologists might encounter, like cellular slime molds (phylum Acrasiomycota and phylum Dictyosteliomycota), plasmodial slime molds (Plasmodiophoromycetes), and the fungal members of the Stramenopila (phyla Oomycota, Hyphochytriomycota, and Labyrinthulomycota). None of these organisms are close relatives of Kingdom Fungi, but several, especially among the Stramenopila, have similar structures and ecological roles to true fungi.

It has been argued that the oldest fossils found to date (which are about 650 million years old) are actually lichens rather than worms or jellyfish; however, this is a hotly disputed interpretation. Intimate associations between fungi and plants occurred very early in evolution. Almost all land plants of today form cooperative associations with fungi, which contribute to the mineral nutrition of the plant and can benefit plants in a variety of other ways. This association is the mutually beneficial mycorrhiza. This cooperation would have eased, if not solved, some of the most difficult problems the first land plants faced as they emerged from the primeval oceans. Some of the oldest (about 400 million year old) plant fossils contain mycorrhizal structures almost identical to those that can be seen today. It is now generally thought that the initial exploitation of dry land by plants about 430 million years ago depended on the establishment of cooperative associations, between fungi and algae on the one hand (as lichens), and between fungi and emerging higher plants (forming mycorrhizas) on the other. Fig. 1.1 compares evolutionary events in fungi with those in some other organisms.

Fungi were crucially important in shaping the ancient ecosystem. An even more radical interpretation, however, is that the oldest terrestrial fossils were actually fungi. The oldest terrestrial fossils we have are made up of masses of threadlike and tubelike structures. They are called *nematophytes*; the name being derived from the Greek *nema*, which means thread, combined with *phyte* because of the belief when originally found that they were plant in origin. Nematophyte fossils start in rocks more than 450 million years old and, in terms of both abundance and diversity, they were important components of the Earth's terrestrial ecosystems for the best part of 100 million years, from the Ordovician to the early Devonian geological periods. They included by far the largest organisms in early terrestrial ecosystems; some specimens of a nematophyte genus called *Protoaxites* have been reported to be more than 1 m wide. These fossils are now being reinterpreted following developments in chemical analysis that suggest that their walls were not composed of the sorts of chemicals you would expect in plant cell walls. As a result, it has been claimed that some of the nematophytes (including *Protoaxites*) were terrestrial *fungi*, thus creating the possibility that the earliest terrestrial organisms were fungi, some of which were far larger than any known today.

Amber is good at preserving soft-bodied organisms like fungi. Fungal spores have been found in amber that is about 220 million years old. Several of these spores are almost identical to fungi that exist today. This is pretty remarkable. When they were trapped in the resin that hardened into amber, all of Earth's land masses were combined into one supercontinent (called Pangaea), birds were only just beginning to evolve, and flowering plants would not appear for another 100 million years! Fossils like this (and others) show that the characteristic fungal structures seen today arose long, long ago and have been maintained for enormous periods of time. One of


common molds
and filamentous
ascomycetes
mushroom
group
yeasts
chytrids
basidiomycetes
ascomycetes
Chytridium
chytrids in
ruminants
Glomus
Ustilago
Tremella
Schizosaccharomyces
Candida
Saccharomyces
Morchella
Neurospora
Ophiostoma
Millions
of
years
Now
100
200
300
400
500
600
Tertiary
Cretaceous
Jurassic
Triassic
Permian
Carboniferous
Devonian
Silurian
Ordovician
Cambrian
Genus Homo appears
Grasses & grazing mammals abundant
Birds, mammals & flowering plants continue to diversify
Dinosaurs become extinct
Asteroid impact?
Major radiation of flowering plants
Dinosaurs abundant
First birds
First mammals
Continents move apart
First dinosaurs
Cycads and conifers abundant
Ice age causes massive extinctions
Mammal-like reptiles appear
Land forms a single continent
Primitive forests widespread
Coal deposits
Reptiles appear
Continents move together
Land plants & arthropods abundant
First amphibians & insects
20% atmospheric oxygen
First fishes with jaws
Animals & plants invade land
Earliest vertebrates appear
2% atmospheric oxygen
Trilobites & brachiopods abundant
All metazoan phyla present
Oldest known metazoans (coelenterates, annelids, arthropods)
1
2
3
4
5
6
7
8
9
10

the experts put it this way: ". . . the history of fungi is not marked by change and extinctions but by conservatism and continuity. . . ." The most remarkable find of amber reported so far is probably that some contain the remains of two mushrooms that can actually be identified because they are so similar to existing mushrooms. The amber, however, is 90–94 million years old. Before the age of mammals, when dinosaurs still ruled the Earth, mushrooms existed that were almost the same as those that exist today. Mushroom fungi first evolved about 200 million years ago, but the mushrooms we see around us when we trek through the forests now are pretty well identical to mushrooms in the undergrowth through which dinosaurs trekked 50 or 100 million years ago. They survived whatever cataclysm brought extinction to the terrible reptiles. They have seen the mammals evolve to a primate that calls himself *sapiens*, and there's no reason to doubt that they'll still be around when all the primates are dead and gone.

If a final reckoning of eukaryotic life on Earth is ever written, the fungi will figure from first to last. They were arguably the first higher organisms to evolve, in a sense giving rise to plants and animals, perhaps 2×10^9 years ago. Later, they enabled plants to invade the land to start terrestrial development of planet Earth, helping the plants to shape nature as we know it today. We would not be here without fungi because their interventions and contributions have been crucial in the development of life on land to the

Fig. 1.1. A cladogram showing phylogeny of the true fungi based on the 18S rDNA gene sequence. Branch lengths in the cladogram are proportional to the average rate of nucleotide substitution (1% per 100 million years), so the cladogram becomes an evolutionary tree, which has been calibrated using fossil fungi, fungal hosts, and/or symbionts. The geological time timescale on the left shows the context of other major evolutionary events in geological time. The numerals and cartoons on the cladogram illustrate major milestones of fungal morphological evolution: terrestrial higher fungi diverged from water molds (1) as branching filaments without septa (2) about 550 million years ago (Mya); the Glomaceae diverged from the progenitor of ascomycetes and basidiomycetes about 490 Mya, and the latter lineage evolved septate filaments (3); clamp connections mark early basidiomycetes (4); basidia (smutlike, 5), asexual spores (6), and asci (7) probably evolved early in the major radiations of basidiomycetes and ascomycetes; filamentous ascomycetes diverged from the yeast lineage about 310 Mya, and fruiting bodies (8) presumably evolved before the Permian divergences because they are present in all the lineages today; mushroom fungi (9), with their characteristic holobasidium (10) probably radiated 130 Mya, soon after flowering plants became an important part of the flora. It is interesting that coals deposited in the Cretaceous and Tertiary periods show much more evidence of fungal decay than the much older Carboniferous coals, which reflects the radiation of aggressive wood decay basidiomycetes from the Triassic onward. (Based on Figs. 5 and 6 in Berbee & Taylor (1993), *Canadian Journal of Botany* 71, 1114–1127.)

point where it could support larger animals. All the while, the fungi themselves were so well adapted to even dramatically changing environments that their own evolution was slow and relaxed. Today, fungi range from among the smallest to the largest individuals on Earth. The yeasts are among the smallest, yet we use them to make enough alcohol every year to refloat the *Titanic*. The largest known organism on Earth is the 890 hectare *Armillaria ostoyae* clone in the Malheur National Forest in Oregon (see Section 2.1).

1.2 Diversity in the Kingdom Fungi: Chytridiomycota, Zygomycota, Ascomycota, and Basidiomycota

The exact number of species in the Kingdom Fungi is unknown, but it is thought to be very large, perhaps as many as 1.5 million, although fewer than 10% of this number have been found and described. In current systematics it is usual to recognize four phyla among the true fungi. In addition, there is a polyphyletic group of imperfect fungi, not given phylum status and called the *Deuteromycetes* or *Deuteromycota*, which largely functions as a "temporary home" for specimens whose true species relationships are yet to be determined. The four phyla of Eumycota or true fungi are the Chytridiomycota, Zygomycota, Ascomycota, and Basidiomycota.

Members of the Chytridiomycota (usually called *chytrids*) may resemble ancestral fungi. The true place for chytrids is among the fungi because their pathway of lysine biosynthesis, their chitinous cell wall, and their small subunit rDNA sequence comparisons show that they and other fungi form a monophyletic Kingdom. Most chytrids live in freshwater or wet soil, but some are marine. They may be saprotrophic or parasitic on organisms such as diatoms, algae, insect larvae, or nematodes. Anaerobic chytrids in ruminant stomachs are responsible for breaking down cellulose for their hosts. This symbiotic assistance in the digestion of plant material by the animal host makes this a mutualistic association, which is crucially important in agriculture and in the evolution of animals. Some chytrids are unicellular; others have a thallus, or mycelium, made up of branching chains of cells that may have filaments (called *rhizoids*) that anchor the thallus to the substrate and absorb nutrients. One of the best-studied chytrids is *Allomyces*.

One of the oldest fossil fungi is a specimen of *Allomyces* from the Devonian period, which was living about 400 million years ago. Today, this genus has an alternation of generations. Haploid zoospores alight on organic material in water and germinate to form small haploid thalli. These produce male and female gametangia in which haploid uninucleate gametes are produced by mitosis. Both gametes have flagella, and the female gamete attracts the swimming male gamete by producing a pheromone. The gametes fuse to form a diploid zygote that gives rise to a small diploid thallus. This produces diploid zoospores that disperse and can germinate to form more diploid thalli. The diploid thallus can also form thick-walled

resting sporangia to survive unfavorable conditions. Meiosis eventually occurs within the resting sporangia so that haploid zoospores are released from it to disperse and start the cycle again. Chytrids are the only true fungi that have flagella at any stage in their life cycle. The phylum Chytridiomycota was the first fungal group to diverge from the others, and the phenotype of the motile stage, which is unicellular with a single posterior flagellum, is shared with the protist group Choanoflagellida, and male gametes of animals. This has given rise to the speculation that choanoflagellates and the fungal and animal kingdoms had the same protist ancestor, which may have looked something like modern-day choanoflagellates.

A key event in the evolution of the fungi after the chytrids diverged from the others was the loss of flagella. By comparing rDNA sequences to establish relationships, flagella have apparently been lost independently several times. This has resulted in the Chytridiomycota and Zygomycota being rather tangled together. Zygomycetes (an informal name for phylum Zygomycota) reproduce sexually by fusion of two gametangia to form a thick-walled zygosporangium that contains zygospores. Zygospores may remain dormant for months before their nuclei undergo meiosis. A sporangium that contains the products of meiosis (sporangiospores) emerges from the zygospores; the sporangiospores disperse and germinate to form a new generation of haploid hyphae. Most zygomycetes have coenocytic hyphae (hyphae without septa, or cross walls). Chytrids are also coenocytic, so both groups are often referred to jointly by the informal name coenomycetes. As with many chytrids, zygomycetes are developmentally indeterminate (i.e., they just keep on growing): however, zygomycetes produce no motile cells, and the zygote is the only diploid cell in their life cycle.

The zygomycete mycelium spreads over its substrate by apical hyphal growth, with occasional stalked sporangiophores reaching up into the air. Almost 900 species of zygomycetes have been described. A commonly encountered zygomycete, both in real life and in textbooks, is *Rhizopus stolonifer*, the black bread mold, and other examples are *Mucor* and *Basidiobolus*. Particularly important groups in the Zygomycotina are the Entomophthorales, most of which are parasites of insects, and the Glomales, which include the arbuscular mycorrhizal fungi that form the most common type of mycorrhizal association with roots of 70% or more of the world's plants. The Glomales is the sister group to the most complex fungi, the Ascomycota together with Basidiomycota, which means that they are each other's closest relatives. The Glomales is also a very old group of fungi because arbuscular mycorrhizas can be detected in the fossil rhizomes of plants that were fossilized about 400 million years ago.

The last two phyla of fungi are the Ascomycota and Basidiomycota. The Ascomycota is the largest phylum of fungi. Approximately half of all described fungal species belong to the ascomycetes (informal name for phylum Ascomycota). The distinguishing morphological character of this group of fungi is a cell in which meiosis occurs (a meiocyte) called an *ascus* (plural asci). An ascus lacks internal septa, and meiospores (called

ascospores) form within it. Hyphae of ascomycetes, on the other hand, are regularly septate, but the septa are incomplete; a pore in each septum permits cytoplasm and organelles, including nuclei, to move readily from one hyphal compartment to the next. There are approximately 30,000 species of ascomycetes already described, and they are placed into two broad groups according to whether the asci are contained within specialized fruiting bodies, more properly called *ascocarps* or *ascomata* (singular = ascoma; see Fig. 1.2). The euascomycetes (or "true ascomycetes") produce an ascoma; the hemiascomycetes ("half ascomycetes") do not.

Hemiascomycetes have a poorly developed mycelium; indeed, most are unicellular. The best known and most extensively studied representatives are the ascomycetous yeasts, particularly baker's or brewer's yeast, *Saccharomyces cerevisiae*, which is probably the most important of all domesticated fungi, as well as the widespread human pathogen, *Candida albicans*. Asexual reproduction is by budding, in which the new (daughter) cell forms as an outgrowth from the old one (the mother cell), or by fission, in which the mother cell elongates and then lays down a septum to divide itself into two daughters. Note, though, that the "classic" fission yeast, *Schizosaccharomyces pombe*, is not closely related to *S. cerevisiae*, and belongs to a group called Archiascomycetes which was the first group of ascomycetes to diverge from the rest, before hemiascomycetes and euascomycetes separated from one another (see Fig. 1.1).

Sexual reproduction in hemiascomycetous yeasts starts with the fusion of two uninucleate haploid cells, usually of opposite mating types (discussed in detail in Chapter 4). A diploid zygote is formed immediately. It can reproduce asexually by budding to form a (uninucleate) diploid cell population in some species; in other species the zygote nucleus undergoes meiosis immediately. Hemiascomycetes have no heterokaryon stage (a stage in which different haploid nuclei co-exist within a cell or hypha). When the diploid nucleus undergoes meiosis, the entire cell becomes an ascus, which will contain four ascospores if spores form immediately after meiosis, or eight ascospores if mitosis occurs after the meiosis. Ascospores germinate to become haploid vegetative cells. The sister group to the hemiascomycetes is the euascomycetes, the members of which are filamentous and produce fruiting bodies. Filamentous growth is an ancestral character; even *S. cerevisiae* can produce a filament of sorts under special conditions. *Galactomyces* and *Dipodascus* are examples of hemiascomycetes that are predominately filamentous.

The euascomycetes, however, include all the characteristically filamentous fungi, which are known as *molds*: organisms like *Aspergillus*, *Penicillium*, and *Neurospora*. Many of the most destructive plant pathogens and the majority of the most severe animal pathogens are euascomycetes. About one third of the euascomycetes live in a symbiotic association with algae in the form of lichens. There are also several groups of mycorrhizal

Fig. 1.2. Line drawings showing construction patterns of some ascomycete fruit bodies (ascomata) (top five), and basidiomycete fruit bodies (basidiomata) in the form of simplified diagrammatic sectional drawings. The black line in each case represents the hymenial tissue (i.e., the tissue layer in which the spore-bearing cells are located). (Fig. 1.2 adapted and redrawn from Moore (1998), *Fungal Morphogenesis*, Cambridge University Press.)

ascomycetes, although the Zygomycota (specifically the Glomales) and Basidiomycota (the Agaricales) are much more important as mycorrhizas. Euascomycetes reproduce asexually by producing spores called *conidia* that form at the tips of specialized hyphae (conidiophores). There is an enormous species-diversity in spore formation and spore morphology. Fossils of many distinctive euascomycete spores have been found in rocks of the lower Cretaceous period, showing that these fungi were already well established about 170 million years ago.

Most euascomycete species are haploid, but nuclear fusion is delayed after mating and the mycelium grows on as a heterokaryon, containing the (haploid) nuclei from both parents. The hyphae also have cross walls, or septa, at regular intervals along their length. Many species make a fleshy fruiting body, which protects the cells that will undergo meiosis during their development, and aids dispersal of the ascospores. In many cases the fruiting bodies are cup shaped with the inner surface of the cup covered with a tissue comprised of a mixture of asci and sterile hyphae, called a *hymenium*. Fruiting bodies are composed of several distinct tissue layers made up of differentiated cells, even though the underlying structure is still that of a filamentous mycelium. Ascus shape varies considerably; long thin asci can forcibly eject their spores, but spores are passively released from globose asci. Fruiting body morphology also varies in relation to spore dispersal and ascomycete systematics has traditionally been based on fruiting body morphology. What seem to be the most primitive fruit bodies have their asci arranged in exposed hymenia that cover parts of open fruiting bodies, called *apothecia*. Other fruiting bodies are flask shaped with narrow openings in the neck (called *perithecia*); some are completely closed (*cleistothecia*) (Fig. 1.2).

Two groups of ascomycetes with unusual fruiting bodies, which are particularly prized as food, are the morels and truffles. Morels produce a large epigeal (i.e., growing on the soil surface) fruit body, which has a stem up to 15-cm tall surmounted by a spongy cap covered in ridges and pits that bear the exposed hymenia. Truffles are mycorrhizal, especially with species of oaks, and their fruit bodies grow underground (= hypogeal). The simplest members of the truffle group produce cup-shaped apothecia, which are almost closed but still have a pore or slitlike opening. The more highly adapted species produce globose fruit bodies that are essentially solid, and have internal chambers and canals lined with hymenium (Fig. 1.2).

The last fungal phylum is the Basidiomycota, which, in evolutionary terms, is a sister group to the Ascomycota. The common morphological character delimitating the group is the basidium, which is a meiocyte produced at the end of a hyphal filament. The diploid nucleus is formed in the basidium and immediately undergoes meiosis. The four resulting haploid nuclei are incorporated into haploid basidiospores that arise externally on tiny stalks. The stalks are called *sterigmata*; the meiospores are called *basidiospores*. These basidiospores typically are forcibly discharged from

their basidia into the air and then germinate, giving rise to haploid hyphae. As an aside, note that the three crown phyla of the fungi have distinctive meiocytes: the basidium in the Basidiomycota, the ascus in the Ascomycota, and the zygosporangium in the Zygomycota.

Life cycles of Basidiomycota typically feature a long phase in which the hyphae are a particular sort of heterokaryon in which the hyphal compartments contain two sexually compatible haploid nuclei, and the (binucleate) hyphal cells are characteristically delimited by septa that have small, central pores with a distinctive organelle (called the *parenthesome*) that protects the pore. This is the dikaryon. Maintaining the dikaryotic state requires that the mitotic divisions of the two haploid nuclei are synchronized. Also, daughter nuclei have to be sorted so that one of each mating type associate together in each daughter cell. In many species (often those with fairly narrow hyphae), a short hyphal branch with a distinctive backwards growth pattern, called either a clamp connection or a hook cell, is coordinated with the synchronized mitosis to manage nuclear sorting (see Chapter 4). A specimen of this structure has been identified in a fossil of a Carboniferous fern that is about 290-million-years old. The hyphal growth pattern that characterizes basidiomycetes (the informal name for the Basidiomycota) was clearly well established that length of time ago.

In all, approximately 25,000 current basidiomycete species have been described. Comparison of ribosomal DNA sequences of basidiomycetes indicates that there are three lineages among them: the Urediniomycetes and Ustilaginomycetes do not have fruiting bodies, but do include the major plant pathogens, and the Hymenomycetes. The latter group makes fruiting bodies, and the basidia are then arranged in a hymenium layer. Many basidiomycetes form mycorrhizas. Many of the Agaricales (generally stalked mushrooms and toadstools) are ectomycorrhizal fungi, meaning the mycorrhizal associations they form with plant roots are characterized by a fungal sheath (Hartig net) which extends both into the root in between the cortical cells and into the soil. Although these ectomycorrhizas are less common than are the endomycorrhizas of the Glomales, they are extremely important for hardwood and coniferous trees in the forests of temperate zones.

Basidiomycetes create the most dramatic and prominent fruit bodies produced by the fungi, including puffballs (which may grow to be more than 50 cm in diameter), mushrooms of many different sorts (some tropical species of which can have caps that also exceed 50 cm in diameter), and the bracket fungi (some of which are perennial) that are often seen on trees and fallen logs in woodland. Many of the latter wood-decay fungi belong to the group called the *Aphyllophorales*. These include both brown rots that specialize in cellulose degradation and white rots that can decompose both cellulose and lignin. Among all the organisms on Earth, the ability to degrade lignin seems to be limited to these fungi.

1.3 Fungi in Nature

There is an enormous fungal diversity extending from polar through to tropical regions, although it is likely that fungi show a similar bias to many other organisms for much greater diversity in the tropics. Fungi exist in a diverse range of habitats. Some are aquatic, and may favor fresh water, brackish areas like mangroves, or marine habitats. Most fungi, though, are terrestrial and live in soil or on dead plant remains; such fungi often make crucial contributions to mineralization and the recycling of nutrients in nature. Many fungi are pathogens of terrestrial plants; indeed, fungi cause the majority of the economically important crop diseases. A few even cause diseases in animals, including humans, although, bacteria and viruses are the most significant pathogens of animals with the exception of immunologically compromised patients.

Traditional systems of fungal classification emphasized morphology above all else. These are now being replaced by schemes that use molecular evidence of relationships and biological species concepts, although the old descriptions still have descriptive value. Fungi that produce large fruit bodies are a good example (Fig. 1.2). The systematic arrangement common at the end of the nineteenth century used the shape and form of the fruit body and especially the nature of the tissue on which the spores were made. Thus, there was a group called *agarics*, which have plates (or gills) beneath an umbrella-shaped cap, just like the ordinary cultivated mushroom. Mushrooms with gills were placed into that group regardless of their other characters. This agaric group was contrasted with fruit bodies that had tubes (or pores) in a spongy layer beneath the cap (called *polypores*). Toadstools in this group were called *boletes*, but *bracket fungi* whose fruit bodies grow directly on the trunks of trees were also included. Then there were those with teeth or spines hanging down below their cap or bracket; these were called *hydnoids*. Other major groups included some with spores formed over the outside of a club-shaped (called *clavarioid*) or corallike (called *coralloid*) fruit body. There were also the completely enclosed fruit bodies (called *gasteroid*) that had their spores inside the fruit body like puffballs.

Work done in the last 25 years of the twentieth century used developmental features and detailed comparisons of anatomy, chemistry, and microscopic characters to reveal natural groupings and evolutionary relationships. By the end of the twentieth century it had become evident that the function of the fungal fruit body is to distribute as many spores as the structure will allow, and its shape and structure reflect that prime functional requirement. The familiar mushroom shape has evolved to give protection to the developing spores. It really is an umbrella protecting the spores from rain. The first step in improving the basic mushroom shape is to expand spore-production capacity. Making gills (platelike downward extensions of the cap) and pores (tubular "excavations" into the cap) are both strategies to increase the surface area available for spore production.

If such tissue structures have positive evolutionary advantages, it is not surprising that the careful observation of developing fruit bodies shows that there are at least 10 different ways by which the mushroom shape can be constructed. It is relatively easy to show that geometrical constraints make pores a less-efficient way of expanding spore production than gills. Thus, there is again no surprise to find that there are some gilled mushrooms that are closely related to polypores and only distantly related to real agarics. Oyster mushrooms are like this. At some stage in their evolution they found advantage in folding their spore-forming tissue into gills and they have now converged onto the agaric shape; it's called *convergent evolution*.

Look closely at the tissue structure and you will find that there are different ways of constructing gills. All mushrooms must increase in size as they develop. Some hold the number of cells unchanged but pump fluids into these cells to increase their size by 10, 20, or 50 times. Others hold the size of cells unchanged and just make more of them to increase the volume of the tissue. Both strategies, though, seem to use the same simple management system whereby one cell organizes and controls a rosette of cells immediately surrounding it. These little families of hyphae orchestrated by a central inducer hypha are called *Reijnders' knots* after the man who first described them.

Look closely at the lifestyle strategies of fungi and you will find some interesting behavior patterns. Bracket fungi achieve massive spore production by increasing the lifetime of the fruit body. At this extreme, the fruit body is a perennial structure. These fruit bodies are often described as woody, but, obviously, because fungi are not plants they cannot use the plantlike wood components. In another example of convergent evolution, they have developed their own solutions to the same challenges solved by woody plants. If fungi are to last several growing seasons they need mechanically strong structures that must be resistant to attacks by pests and microbes, as well as to adverse weather conditions.

At the other extreme of the lifetime strategy spectrum, there are some fungi whose fruit bodies last barely more than a day. They may mature overnight and be dead and gone by the next night. These mushrooms are stripped down for athletic action. They tend to be small and delicate and adapted to get maximum spore yield from minimum mass of fruit body.

Stinkhorns are interesting because they parallel mushrooms in gross morphology; they have recognizable caps and stems. The whole structure, however, is adapted to insect dispersal, as opposed to wind dispersal, of the spores. Like insect-attracting flowering plants, these fungal fruit bodies sport bizarre shapes, colors, and penetrating smells to attract flies and other insects. A distinctive odor is also important for fruit bodies formed under the soil surface. Truffles are the obvious examples, but other fungi use the same strategy of attracting animals (some produce an odor resembling the male sex hormone of pigs) to dig them up and help disperse their spores.

Fungi make a crucial contribution to the present day ecosystem in a wide variety of ways. Almost all fungi exist primarily as mycelium; the only numerous exceptions are the yeasts. Fungi and bacteria help to maintain the balance of nature by nutrient cycling. It is conventional to group microorganisms into metabolic classes, depending on the sources of nutrition they use. All the terms used to describe these classes employ the combining form *troph,* derived from a Greek word meaning "to feed." Fungi decompose things, and decomposers are called *saprotrophs*. Organisms that use light as an energy source are called *phototrophs* (*photo* is from the Greek word for "light"); organisms that use chemicals as energy sources are called *chemotrophs*. Those that use organic compounds as energy sources are types of chemotrophs called *chemoorganotrophs*, whereas organisms able to use inorganic chemicals as energy sources are called *chemolithotrophs*. A classification by ecological nature identifies saprotrophs, for which dead organic materials not killed by the organism itself provide the nutrients; *necrotrophs*, which invade living tissues that they kill and then utilize; and *biotrophs*, which exploit host cells that remain alive. Biotrophs may be host-specific, but saprotrophs and necrotrophs generally have a very large range of habitats and nutrient sources open to them.

A particularly important fact about fungi is that they are the only organisms that can digest wood because the lignin, which is complexed with the cellulose in wood, is so difficult to degrade. Lignin digestion is a fungal specialty, but they gain nutrition from living or dead animals as well as plants, and the breakdown of other complex molecules such as cellulose and tannins in soils is due mostly to fungal enzymatic activity. The total length of hyphae in grassland soil has been estimated at more than $1\,km\,g^{-1}$. Fungi make up 90% of the total living biomass in forest soils.

Because they can digest and extract nutrients from so many of the materials that exist on, within, and under the soil, fungal mycelia act as sinks of organic carbon and nitrogen in the soil. In many forests, a lot of carbon fixed by photosynthesis ends up in fungal mycelium because of the mycorrhizal symbiotic association between fungus and forest tree roots. It is important that the mycelial sink keeps the nutrients on site preventing loss from the soil by leaching. The benefits of mycorrhizas to plants include efficient nutrient uptake, especially phosphorus, enhanced resistance to drought stress, and direct or indirect protection against some pathogens. Mycorrhizal fungi also link plants together into communities that are more resilient to stress and disturbance than single plants. When a common fungus connects plants, the products of photosynthesis can move through the fungus from a well-placed donor plant to a shaded recipient plant. In nutrient-poor soils, mycorrhizal fungi can provide nitrogen to their host plant that their mycelia have obtained by saprotrophic digestion of nutrients in the soil. The mycorrhizal interconnections form a network through which plant-to-plant, plant-to-fungus, and fungus-to-plant transfers of nutrients can take place.

Soil fungi also exude polysaccharides during mycelial growth that are important in gluing soil particles into aggregates that improve soil aeration and drainage. This is a critical contribution the fungi make to soil structure because most terrestrial organisms are so strongly aerobic.

There is such a large amount of fungal mycelium in most soils that it makes a major contribution to food webs by being eaten by numerous vertebrates and invertebrates, including insects, mites, mollusks, and nematodes. Microarthropods are responsible for shredding organic matter in soil (and so prepare it for the final mineralization processes carried out by microbes), but about 80% of the tens of thousands of microarthropod species in forest soils are fungivores. When the mycelium makes fruit bodies (e.g., mushrooms, bracket fungi on tree trunks, and truffles), these are also vital food sources for many animals, from mollusks to mammals (including humans). Because most small mammals depend on underground (hypogeous) fruit bodies, especially truffles, for a significant part of their diets, the influence of fungi on the food web extends even further. Top predators (e.g., birds of prey) that eat small mammals that subsist almost entirely on hypogeous fungal fruit bodies, are themselves indirectly dependent on the fungi of the soil.

Even fungal plant pathogens have a positive effect by enriching the structure of the natural environment. Plants killed by disease provide organic matter for nutrient cycling. Dead branches or heart rot in live trees create habitat for cavity nesting animals, whereas gaps in stands of dominant plants created by disease allow development of other plants, which contributes both to species diversity and diversity of food for animals, from insects to elk.

1.4 Fungi in Technology

Humans have been using fungi for many thousands of years. More than 5000 years ago a Neolithic traveler set out across the Alps. He died in the ice and snow, and his corpse was preserved in the glacier, eventually to be exposed at the edge of the ice sheet in 1991 close to the Austrian–Italian border. A remarkable feature of the Iceman's equipment is that it included three separate fungal products. One was a mass of fibrous material of a fungus with a long history of use as tinder, so it was clearly part of the Iceman's fire-making kit. The other two were pieces of a bracket fungus threaded onto leather thongs. The fungus (actually called *Piptoporus betulinus*) is known to accumulate antiseptics and chemicals (known as *secondary metabolites*) able to reduce fatigue and sooth the mind. Perhaps they were chewed on when the going got tough on his trek over the Alps.

Our distant European ancestors evidently held fungal products of this sort in such high esteem that they were necessary accessories for hazardous journeys. Today, alcohol and citric acid are the world's most important

fungal metabolites in terms of production volume, although penicillin is still an important antibiotic. Most antibiotics that we use today actually originate from bacteria, but fungi produce some other useful pharmaceuticals (e.g., compounds capable of suppressing the immune response in transplant patients so as to avoid organ rejection like the fungal product called *cyclosporin*). Another natural compound obtained from fungi that has great medical value is called *mevinolin*. This is produced by the fungus *Aspergillus terreus* and is the basis of the "statins," which are used to reduce cholesterol levels (high cholesterol levels being considered to be a risk factor in heart disease). Three compounds derived from mevinolin—Pravastatin, Simvastatin, and Lovastatin—had worldwide sales that put them individually in the top 10 selling pharmaceuticals, with combined sales valued at $5 billion in the late 1990s.

Fungi also produce compounds like the ergot alkaloids, steroid derivatives, and antitumor agents. Ergot toxins in low, controlled, concentrations are valuable drugs that cause vasodilation and a decrease in blood pressure, as well as contraction of smooth muscles. Although the alkaloids can be synthesized, strain improvement by mutation and selection of high-yielding strains have been so successful that fermentation remains the most cost-effective means of production. Most of the steroids in clinical use today are modified by fungi during manufacture. By using fungi to make specific chemical transformations, compounds can be made that would otherwise be very difficult, impossible, or just too expensive to produce by direct chemical synthesis. Many fungal products have recently been found to inhibit the growth of cancers in animal tests. Specificity and safety are the issues that presently limit the medical usefulness of most of these compounds (they may have adverse effects on healthy tissue as well as the tumor). Compounds that work by modifying the activity of the patient's own immune system (immunomodulators), making it more active against the cancer cells, seem likely to offer the greatest value.

Mushroom cultivation is the most obvious way in which we exploit fungi. In this worldwide industry about 30 species are commercially cultivated, seven being cultivated on what could be described as industrial scale. Mushroom cultivation, however, suits all scales from peasant farms through smallholdings and on to multimillion dollar, highly mechanized farms.

In agriculture, fungal pathogens may be useful for the biological control of weeds and other pests. In the United States, 13 of the top 15 weeds are alien species; in Canada, 78 of the 107 most troublesome weeds were introduced to the country. The best way to control such alien plants is to bring in diseases from their home territories. This policy has an overall success rate of 67%. Fungal diseases could be used as control agents of other crop pests, like insects, mites, and nematode worms. Biological control strategies have the advantage of being particularly highly specific and offer freedom from excessive use of chemical pesticides; however, the concept of biocontrol remains promising but largely unproven.

The ability of fungi to recycle just about any material brings them to the forefront for tackling current problems of waste and pollution. Biotechnological alternatives to fossil fuels might include materials like alcohol produced, for example, by fermentation of industrial waste products such as molasses, straw, and wood pulp. Although these are expensive alternatives at the moment, such biofuels are likely to become viable alternatives as increasing taxes are imposed on oil-derived fuels following international agreements based on the principle that the polluter pays. One thing we must do is to make better use of what the Earth already produces. On average, agriculture currently loses 40% of its primary production to pests and diseases, and then throws away more than 70% of what's left because the crop always represents so little of what is grown. That works out to an overall 18% efficiency, at best, for the world's agriculture. Fungi, especially the mushroom fungi, are ideal candidates to degrade waste vegetation, including any polluted with pesticides. Many pesticides are chemically similar to the complex phenolic compounds found in wood, and because the fungi can decompose wood, they can also be used to degrade environmental pollutants in soils and wastewater discharges.

1.5 Fungi as Models

For 1000 million years (a quarter of the time there's been life on the planet), the eukaryotic organisms were not recognizable as animals, fungi, or plants, but were some sort of primitive unicellular ancestor of all three. Today, the three main eukaryotic Kingdoms (animals, fungi, and plants) are the only ones to have evolved complex multicellular development, an ability that emerged about 600 million years ago. Most of the cellular machinery that is necessary for development appears early in eukaryotic evolution, and is expressed by many lineages. This machinery includes a common system of gene regulation and structure, the ability of cells, and localized regions of cells, to differentiate to specific functions, the main cellular organelles, the cytoskeletal architecture, signal transduction mechanisms, and cell-to-cell signaling. Because these components of the eukaryotic cell occurred so early in evolution, animals, fungi, and plants alike share them all.

Because so many fungi are easy to grow and manipulate in the laboratory, however, several fungi have been developed into model organisms that are studied because they are representative of all other eukaryotes. The prime example of such a model fungus is undoubtedly the yeast *Saccharomyces cerevisiae*.

The word *yeast* is a general term for any growth that appears in a fermenting liquid. In its origins, the word means frothy or foamy. Thus, it is descriptive of the fermentation process, but it has become associated with the agent of fermentation. When grape juice is collected it ferments, and the growth that occurs and eventually forms sediment is "yeast." Making

alcoholic drinks is such a simple process that all, even the most primitive, societies have one or more fermentations that they include in their rituals. There are some ancient Egyptian murals and tomb ornaments that depict both bread and winemaking. From the biological point of view it is remarkable that the one organism responsible for most fermentations is the yeast now called *Saccharomyces cerevisiae*, but known as *Brewer's yeast* in one trade and *Baker's yeast* in the other.

At the end of the nineteenth century, industrialization created the need to guarantee and improve production and product quality, and this prompted brewers and winemakers to sponsor research into the nature of fermentation. This naturally came to focus on the single-celled microorganisms we call yeast. Studies of yeast metabolism essentially founded the sciences of biochemistry and enzymology. Purification of cultures (necessary for a uniform product) and a drive to improve cultures (to increase the efficiency of fermentation or develop new products) was enhanced by the parallel development of the science of genetics at the beginning of the twentieth century. At about the time that Mendel's work on garden peas was being rediscovered and republished, it was becoming evident that vegetative cells of *S. cerevisiae* are usually diploid, produced by "copulation" of two haploid spores. In the early 1930s the basic facts of the yeast life cycle were established. Single spores planted in culture medium germinated to produce a cluster of round haploid cells. The large, ellipsoidal vegetative yeast cell is produced by the fusion of two round haploids. It was found that this species possesses two mating types in the haplophase, and that haploid cells of different mating types mate to produce the diploid. The diploid nuclei subsequently undergo meiosis (the reduction division), during spore formation, to produce four haploid ascospores.

With the life cycle clearly established, the way was open for breeding experiments. By the 1940s, it was possible for Carl Lindegren to write:

> Thirteen asci were analyzed from a heterozygous hybrid made by mating a galactose fermenter by a nonfermenter; two spores in each of these asci carried the dominant gene controlling fermentation of galactose, and two carried the recessive allele. A backcross of fermenter [offspring] to the fermenter parent produced 13 asci; all four spores in each of these asci carried the fermenting gene. A backcross of the nonfermenter to the nonfermenting parent produced seven asci, each of which contained four nonfermenting spores. A heterozygous zygote was produced by backcrossing a nonfermenter [offspring] to the fermenting parent; six asci were analyzed and each contained two nonfermenting spores. This analysis shows quite convincingly that the genes controlling fermentation of galactose behave in a regular Mendelian manner.

Within 50 years of the rediscovery of Mendel's experiments with peas, therefore, those experiments had been repeated with yeast and had demonstrated that yeast genes operated to the same set of rules as did pea genes. Within another 50 years the yeast cell was regularly being used in experi-

ments involving gene cloning and sophisticated genetic modification. Indeed, in April 1996 the DNA sequence of the complete genome of *S. cerevisiae*, the first eukaryote genome to be sequenced, was published. This was the result of the combined efforts of 100 European laboratories.

1.6 Genes to Genomics

The genome is made up of the entire DNA content of a cell. Eukaryotes and prokaryotes have quite different types of genome, but we must assume that the prokaryotic grade of organization is the primitive form from which the eukaryote organization evolved. They have a great deal in common; the DNA of a gene is transcribed into RNA, which is called a messenger RNA (mRNA) if it is a transcript of a protein-coding gene. The mRNA is translated into protein by the ribosomes and other translation machinery, that part of a protein-coding gene sequence that is translated into protein is called the *open reading frame*, usually abbreviated to ORF. The ORF is characteristically read in the 5′ to 3′ direction along the mRNA, and it starts with an initiation codon and ends with a termination codon. Nucleotide sequences that occur in the mRNA before the ORF make up the leader sequence, and sequences following the ORF make up the trailer segment. Many eukaryotic genes are split into *exons* (meaningful segments) and *introns* (sequence segments that do not contribute to the protein-coding sequence). The introns are removed from the primary RNA transcript by the splicing machinery to form the functional mRNA.

Prokaryotic genomes are generally much smaller than those of eukaryotes. The *Escherichia coli* genome, for example, is composed of 4.64 Mbp (megabase pairs) of DNA, that of *Streptomyces coelicolor* is 8 Mbp, whereas the yeast genome is 12.1 Mbp and the human genome is 3000 Mbp. The physical organization is also different because the genome in prokaryotes is contained in a single, circular, DNA molecule. Eukaryotic nuclear genomes are divided into linear DNA molecules, each contained in a different chromosome. In addition, all eukaryotes have mitochondria, and these possess small, usually circular, mitochondrial genomes. Photosynthetic eukaryotes (i.e., plants, algae, some protists) have a third small genome in their chloroplasts.

The size range of the genome corresponds to some extent with the degree of complexity of the organism, but the fit is not exact by any means because this correlation depends on the structure and organization of the genes. For example, the *Escherichia coli* genome has 4397 genes and the yeast genome comprises about 5800 genes, so you might feel confident about believing that yeast has more genes because it is a eukaryote, and you can understand why it doesn't have *many* more because it's a fairly simple eukaryote. The genome of the streptomycete bacterium *Streptomyces coelicolor*, however, contains more than 7000 genes. This organism is a prokaryote, but it has

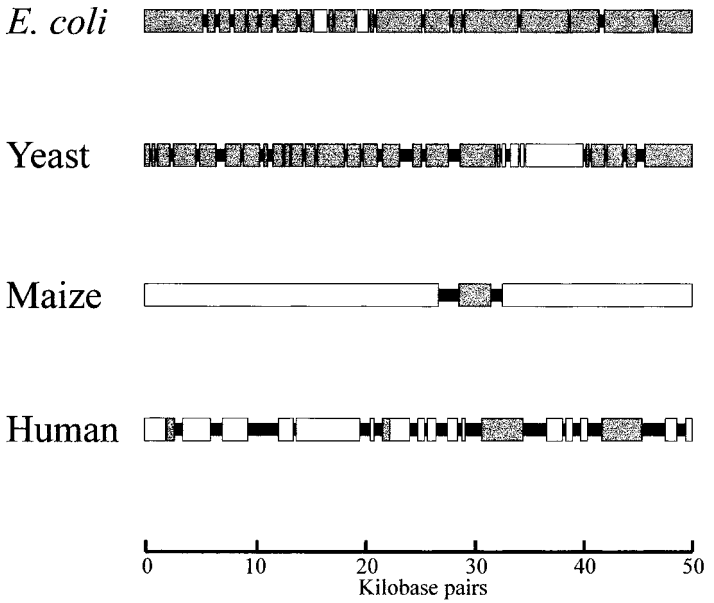


Fig. 1.3. Comparison of 50 kbp segments of the genomes of four organisms to show how the "density" of genetic information varies. In each case the gray boxes correspond to gene sequences, and the white boxes correspond to stretches of repeated sequences. (Fig. 1.3 adapted and redrawn from Figure 1.7 in Brown (1999), *Genomes*, BIOS Scientific Publishers.)

nearly 30 percent more genes than the model eukaryote, yeast. *Streptomyces* is admittedly a fairly complex bacterium and highly advanced in an evolutionary sense; but it *is* a bacterium. The arithmetic difference lies in the fact that the average yeast gene is 2200 base pairs long, whereas the average *S. coelicolor* gene is only 1200 base pairs long; however, we can't explain *why* such a difference in gene size exists. The smallest eukaryotic genomes (like yeast), however, are in the region of 10 Mbp, and the largest are more than 100,000 Mbp (in vertebrates and plants), so we can observe even more surprising structural differences when we compare other eukaryotes. In genera, it appears that space is saved in the genomes of less complex organisms by having the genes more closely packed together and by having much less repetition (Fig. 1.3). The genome of *Saccharomyces cerevisiae* contains more genes per unit length of DNA than occur in human or maize DNA. Also, very few yeast genes are interrupted by introns. In fact, there are only 239 introns in the entire yeast genome whereas some individual genes of higher eukaryotes can have many introns. For example, in humans muscular dystrophy is caused by a lesion in a gene on the X-chromosome, which contains 80 introns; in this case the gene sequence is 2.3 Mbp long, but the mRNA is only 14 kb long, so only 1% of the gene is found in the mature messenger.

Finally, there are fewer genome-wide repeats in the yeast genome; only 3.4% of the total DNA of yeast is taken up by repeated sequences. Compare this with the maize genome: this consists of 5000 Mbp (larger than the human genome but still relatively small for a plant), but noncoding repetitive elements are estimated to make up approximately 50% of the maize genome.

Yeast is a perfectly respectable eukaryote that features all the basic cell biology expected of this grade of organization, but its genome is only in the same size range as some of the more advanced prokaryotic genomes despite the major differences between prokaryote and eukaryote cell structure. This small genome is one reason why yeast geneticists and molecular biologists pioneered eukaryote genome analysis. Although some of the more unusual aspects of genome structure observed in higher animals and plants might not be represented in fungi, the genomes of yeast and other fungi remain good models of eukaryotic genetic architecture. Their smaller size means that the information they contain is technically more accessible. In terms of genetic information content the organization of the fungal genome is much more economical than that of higher eukaryotes. Genes are more compact with fewer introns, spaces between genes are short, and much less of the DNA is devoted to repetitive noncoding sequences. Nevertheless, the genetic structure and functioning of fungal genes is representative of all eukaryotes and we can use their sequences to learn about genomics.

Analysis of the sequences contained in the genome of an organism and comparison between the genomes of different organisms (exercises that have come to be known as the science of *genomics*) only became possible from the mid-1990s. The first complete bacterial genome sequence (of the pathogen *Haemophilus influenzae*) was published in July 1995. This was followed by a number of other bacterial genomes and, as we've stated, the first eukaryotic (the yeast) genome appeared in April 1996. At the moment the sequence databases contain relatively few full genome sequences, but some illuminating comparisons can already be made. For example, it appears that obligate pathogenic bacteria have dispensed with many genes involved in specifying enzymes and regulation of biosynthetic metabolism. It is interesting that the larger genome of *Streptomyces coelicolor* as compared with *Escherichia coli*, for example, can largely be accounted for by a multiplicity of regulatory genes, particularly ATP-binding cassette transporters and two-component regulators, that might be inferred to be needed for the more complex developmental processes related to the pseudohyphal life style of the streptomycete.

Getting the exact DNA sequence of a genome is the prelude to an intensive analysis in which the first step is probably to look for potential start and stop codons. This identifies a collection of potential ORFs, which are the potential genes, contained in the sequence. These are then usually compared with known sequences in database searches because a strong match to a gene known from another species is the clearest way of establishing

that a gene exists as a protein-coding entity. All of this requires sensible use of sophisticated computer algorithms and good computational assistance. The process is often called "data mining" and it depends on what has become a new science, called *bioinformatics*, which assembles, documents, maintains, and analyzes DNA sequences by the megabase.

The ability to sequence and compare complete genomes promises to improve our understanding of many areas of biology. Such data will more directly reveal evolutionary relationships, and indicate how pathogens spread and cause disease. They should enable us to get a comprehensive understanding of the activities of living cells and how they are controlled at the molecular level. The information has practical value, too, which is why so many pharmaceutical companies are involved in genome projects. The hope is that by knowing the genes in such great detail it will be possible to identify those genes responsible for, or which have influence on, diseases, and then design therapies to combat the disease.

In the rest of this book, we will describe the information about basic fungal genetics that contributes to the establishment of genome sequences and subsequently to their interpretation.

Publications and Websites Worth a Visit

Bennett, J.W. (1998). Mycotechnology: the role of fungi in biotechnology. *Journal of Biotechnology* **66**, 101–107.

Berbee, M.L. & Taylor, J.W. (1993). Dating the evolutionary radiations of the true fungi. *Canadian Journal of Botany* **71**, 1114–1127.

Berbee, M.L. & Taylor, J.W. (1999). Fungal phylogeny. In *Molecular Fungal Biology* (R.P. Oliver & M. Schweizer, eds.), pp. 21–77. Cambridge University Press: Cambridge, U.K.

Bos, C.J. (1996). *Fungal Genetics: Principles and Practice*. Marcel Dekker, Inc.: New York.

Brown, T.A. (1999). *Genomes*. BIOS Scientific Publishers: Oxford, U.K.

Carlile, M.J., Watkinson, S.C. & Gooday, G.W. (2001). *The Fungi*, second edition. Academic Press: London.

Gow, N.A.R. & Gadd, G.M. (1995). *The Growing Fungus*. Chapman & Hall: London, U.K.

Hardham, A.R. (2001). Investigation of Oomycete cell biology. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 127–155. Oxford University Press: Oxford, U.K.

Lewin, B. (2000). *Genes VII*, seventh edition. Oxford University Press: Oxford, U.K.

Kendrick, B. (2001). *The Fifth Kingdom*, third edition. Focus Texts & Mycologue Publications, Canada. *The Fifth Kingdom* online at http://www.mycolog.com

Momany, M. (2001). Cell biology of the duplication cycle in fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 119–125. Oxford University Press: Oxford, U.K.

Moore, D. (2000). *Slayers, Saviors, Servants, and Sex: An Exposé of Kingdom Fungi*. Springer-Verlag, New York (a book for the general reader about the world of fungi).

O'Connell, R.J. & Carzaniga, R. (2001). Electron microscopy of filamentous fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 175–208. Oxford University Press: Oxford, U.K.

Oliver, R.P. & Schweizer, M. (1999). *Molecular Fungal Biology*. Cambridge University Press, Cambridge, U.K.

Sugiyama, J. (1998). Relatedness, phylogeny, and evolution of the fungi. *Mycoscience* **39**, 487–511.

Talbot, N.J. (2001). *Molecular and Cellular Biology of Filamentous Fungi*. Oxford University Press: Oxford, U.K.

Thornton, C.R. (2001). Immunological methods for fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 227–257. Oxford University Press: Oxford, U.K.

von Döhren, H. (2001). Biochemical methods for filamentous fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 209–226. Oxford University Press: Oxford, U.K.

Wainwright, M. (1992). *An Introduction to Fungal Biotechnology.* John Wiley & Sons: Chichester, U.K.

Historical Publications Worth Knowing About

Esser, K. & Kuenen, R. (1967). *Genetics of Fungi*. Springer-Verlag: New York.

Fincham, J.R.S., Day, P. R & Radford, A. (1979). *Fungal Genetics,* fourth edition. Blackwell Scientific Publications: Oxford, U.K.

Lindegren, C.C. (1949). *The Yeast Cell, its Genetics and Cytology*. Educational Publishers, Inc.: St Louis.

CHAPTER 2

Genome Interactions

Revision Concepts

- Completely asexual fungi, called *deuteromycetes*, generate variation by modifying genetic expression or undergoing mitotic recombination.
- Fungal sexuality is displayed as homothallism (individuals are self-fertile) or heterothallism (individuals are self-sterile).
- Hyphae with two or more kinds of nuclei are heterokaryotic, whereas hyphae with only one kind of nucleus are homokaryotic.
- A monokaryon is a homokaryotic mycelium with uninucleate cells.
- Hyphal anastomosis is hyphal fusion, followed by nuclear migration between mycelia on condition that they are vegetatively compatible.
- Oomycete and zygomycete hyphae do not anastomose and generally only form septa to partition reproductive or injured structures.
- Specialized sexual structures and pheromones to mutually attract these structures are required in oomycete and zygomycete reproduction.
- Hyphal anastomoses occur freely in higher fungi.
- In yeasts, fusion between individual cells is called *conjugation.*
- Vegetative compatibility genes regulate hyphal fusion, whereas mating-type factors regulate the ability of nuclei to undergo karyogamy and meiosis.
- Vegetative compatibility is governed by *het* (for ***het***erokaryon) or *vic* (for ***v***egetative ***i******c***ompatibility) genes.
- Colonies with the same vegetative incompatibility alleles complete hyphal anastomosis, thus allowing nuclear migration to give rise to a heterokaryon.
- In incompatible reactions fusion cells are sealed off and killed.
- Unifactorial incompatibility involves mating between mycelia that have different alleles at the single mating-type locus.
- Bipolar heterothallism is characterized by equal numbers of progeny of two mating types.
- Mating-type factors are called *idiomorphs* rather than alleles because of their very different lengths and DNA sequence homologies.

- Many basidiomycetes have two unlinked mating-type factors (bifactorial incompatibility); a compatible interaction occurs between mycelia with different idiomorphs at both mating-type loci.
- Tetrapolar heterothallism occurs when meiosis generates progeny of four different mating types.
- Primary homothallism occurs in fungi completely lacking heterothallism.
- Secondary homothallism results from spores with more than one post-meiotic nucleus producing a heterokaryotic but self-fertile mycelium.
- Separate genes for compatibility and mating types occur in ascomycetes.
- Ascomycetes have only two mating types specified by a single locus with different idiomorphs.
- Mating in *Saccharomyces cerevisiae* is controlled by *MAT*, a complex genetic locus comprising two linked genes.
- Yeast mating-type factors specify peptide hormones, which organize the mating process, and corresponding receptors specific for each pheromone.
- *Neurospora* species exhibit different mating strategies: bipolar heterothallism with mating types *A* and *a*, secondary homothallism and two types of primary homothallism.
- Mating results in asci that contain the products of a single meiosis.
- In filamentous ascomycetes, a postmeiotic mitotic event results in each ascus containing an octad comprised of four pairs of sister ascospores.
- In basidiomycetes, mating-type factors regulate self–nonself recognition, mycelial morphogenesis, karyogamy, meiosis, and multicellular development.
- Growth of the basidiomycete dikaryon requires that the two nuclei complete mitosis together (conjugate division) using a mechanism of nuclear migration that depends on formation of a hook cell at each septum.
- In basidiomycetes, mating-type factors, *A* and *B*, regulate sexual reproduction.
- The *A* factor is responsible for conjugate division and formation of clamp connections, whereas the *B* factor is responsible for nuclear migration and clamp cell fusion; a compatible mating requires that both *A* and *B* are different.
- *Ustilago maydis*, the smut disease of maize, has a tetrapolar mating system.
- *U. maydis* sporidia are nonpathogenic and grow vegetatively by budding like a yeast phase; the dikaryon, which grows as a filamentous pathogenic fungus, is produced when sporidia of opposite mating type fuse.
- Fusion of sporidia is controlled by the biallelic *a* mating-type locus, whereas the multiallelic *b* locus determines true hyphal growth and pathogenicity.
- Proteins encoded by the *b* locus contain sequence homology with DNA-binding homeodomain regions of known transcription-regulating factors.
- The homeodomain is a DNA-binding motif encoded by a conserved DNA sequence called the *homeobox*, which is particularly associated

with transcriptional regulators of *Hox* genes that orchestrate development in higher eukaryotes.
- In *Coprinus cinereus*, *Schizophyllum commune* and other basidiomycetes, *A* genes encode transcription factor homeodomain proteins and *B* genes encode lipopeptide pheromones and pheromone receptors.
- No intragenic recombination occurs within the *A* locus because loci are organized into cassettes and are transcribed in opposite directions.
- Homeodomain genes, called HD1 and HD2, at *A* are functionally independent and redundant because only one compatible HD1/HD2 gene combination is required to promote sexual development.
- The multiallelic *B* mating-type factor codes for several pheromone and receptor genes that control *B*-regulated nuclear migration and clamp cell fusion.
- Mating-type factors may not regulate events beyond the mating reaction (e.g., fruiting).
- Sexual reproduction produces genetic variation, thus enhancing adaptation of the species, offering an escape from DNA parasites, and providing a means to repair DNA damage.

2.1 Fungal Lifestyles: Hyphal Fusions Are the Key to Advanced Hyphal Systems

Fungi differ from most other eukaryotes in that their vegetative body (i.e., the thing that grows as an individual and gives them body mass) shows indeterminate growth. Fungal mycelia will continue to grow and invade new substrates for as long as new substrates remain available and growth conditions remain satisfactory. You could take that statement to mean that fungi would just keep on growing until they became enormous, and that does happen in some cases. The largest known organism on Earth is a tree root pathogen known as *Armillaria ostoyae*, a clone of which covers an area of 890 hectares in the Malheur National Forest in eastern Oregon. It weighs in at around 150 metric tons and is at least 2400 years old. The unusual life style of *Armillaria*, however, enables it to get so large; it forms rhizomorphs that scavenge widely through the soil for nutrients and hosts. It is probably true to say that the majority of fungi are "microfungi" that live in very small habitats, like individual leaves, individual insects, or even smaller grains of soil. Like the extent of the habitat the rate at which the fungus can grow will also limit the extent of growth that can occur. For example, lichens can grow in some very extreme environments, from deserts to the Arctic wastes. In the Arctic, however, lichen growth is around 5 cm in 1000 years, so colonies several thousand years old will be little more than the size of your hand. Nevertheless, unless other growth conditions impose a limitation, a fungus will continue to grow until it occupies all of the available substrate. Animals and plants cannot manage that.

The fungal kingdom is very large and diverse. As a consequence of this the part played by sexual reproduction in the life cycle of fungi is also very diverse. There are fungi at one extreme of the spectrum of behavior that are completely asexual organisms (most examples are usually classified among the group known as the *deuteromycetes*, or, more formally, the *Deuteromycotina*). These organisms are not static in an evolutionary sense because they can generate variation by modifying genetic expression or adapt processes that occur during the mitotic division. This can produce asexual propagules in which chromosomes have segregated in new combinations or which contain recombinant chromosomes (see Section 5.7). At the other extreme, fungi display many forms of sexuality that govern the bringing together of genetic information from different parents into some arrangement that eventually produces a (potentially heterozygous) diploid nucleus. This nucleus undergoes meiotic division during which chromosomal segregation and genetic recombination take place as they do in every other eukaryote.

We have to define at least some of our vocabulary before we can take the discussion very much further. A species is *homothallic* if an individual can complete the sexual cycle on its own, but, as we will explain later, there are different ways in which homothallism can be achieved. We should also emphasize that a homothallic species is not limited to self-fertilization. Two homothallic strains may well interbreed, either in nature or with "assistance" in the laboratory. The point is that a homothallic species does have the ability to self-fertilize. This contrasts with a *heterothallic* species, which requires interaction of two different individuals to complete the sexual cycle. Individual isolates of heterothallic fungi are self-sterile or self-incompatible, but they can be cross-compatible. Heterokaryosis results from the fusion of hyphae of different isolates, followed by migration of nuclei from one hypha into the other, so that the hyphae come to have two kinds of nuclei. Such a hypha is a heterokaryon. If there is only one kind of nucleus in the hypha, we have a homokaryon. In the most highly adapted version of this behavior, in model basidiomycetes like *Coprinus cinereus* and *Schizophyllum commune*, a basidiospore germinates to produce homokaryotic mycelium with uninucleate cells, called a *monokaryon*. When two monokaryons meet, hyphal anastomoses occur. If these are vegetatively compatible, nuclei of one migrate into the mycelium of the other. In addition, if the nuclei have compatible mating types, the new growth and the cells of much of the pre-existing mycelium now have binucleate cells, with one nucleus of each monokaryotic parent. This mycelium is called a *dikaryon*.

If a sexual cycle is really important as a means of providing variation, the mechanisms that enable it to occur presumably evolved at a very early stage in evolution. To get an idea of how the mechanisms that organize sexual reproduction might have arisen, we can look at some members of the most ancient groups of fungi and fungal relatives. Analysis of molecular data

indicates that all true fungi have a common origin, and there are some organisms that are relatives with structures and ecological roles similar to true fungi (see Section 9.1). Phylum Oomycota (less formally known as the oomycetes) is among the latter and includes a number of important species with interesting sexual cycles. Members of phylum Zygomycota, which is a phylum of the true Kingdom Fungi, reproduce sexually by fusion of two gametangia to form a thick-walled zygosporangium containing zygospores. The way they arrange this is also instructive.

Vegetative hyphae of the oomycetes are diploid and lack cross walls, except where reproductive organs or damaged parts of the mycelium are separated off. Though heterokaryons of *Phytophthora megasperma* have been forced by using complementing mutants *in vitro* and in plant tissue, a key feature is that hyphal anastomosis is rare except between specialized sexual organs. Sexual reproduction involves a female structure called the *oogonium*, in which one or several uninucleate oospheres differentiate. When these are fertilized they give rise to thick-walled oospores. The fertilizing nuclei come from a male structure (antheridium) that develops alongside the oogonium. Meiosis occurs in the oogonia and antheridia; the postmeiotic haploid nuclei being packaged into gametes. Fusion of one male and one female nucleus occurs in the oospore.

The genus *Achlya* has been used to study oomycete sexual reproduction. Self-sterile *Achlya* strains that reproduce only when paired with another strain are frequently found, and these can be described as being heterothallic. The population is not neatly differentiated into homothallic and heterothallic strains, however, but rather into strains with varied sexual potency. Some strains are predominantly males, acting as antheridial parents in crosses, or predominantly females, acting only as oogonial parents; still other strains are intermediate, acting as males in crosses with predominant females or as females when crossed with a predominant male. This is called *relative sexuality*.

The first step in the reproductive cycle is that the female produces a hormone (a steroid called *antheridiol*) that induces the male strain to produce antheridial branches. The induced male strain then produces a second hormone (another steroid called *oogoniol*), which induces the female to form oogonial initials. There are specific receptor proteins in each sex for the hormones of the other, and activation of the receptors leads to acetylation of histone proteins, activation of specific mRNA synthesis, and consequential synthesis of specific proteins. The genus *Phytophthora* includes both homothallic and heterothallic species. For example, *P. cactorum*, *P. heveae*, and *P. erythroseptica* are homothallic, and *P. infestans*, *P. palmivora*, and *P. cinnamomi* are heterothallic. In the heterothallic species there are two mating types, called *A1* and *A2*, which are both capable of forming antheridia and oogonia, although neither of them can produce oospores when alone, only when paired. This involves two hormones. The one produced by the A1 mating type that induces A2 is called α 1 hormone,

and that produced by A2 to induce A1 is called α 2. The chemical nature of *Phytophthora* hormones is not yet known. It has been shown that *Phytophthora* and *Pythium* species require sterol supplements for sexual reproduction, as well as production of normal sporangia and motile zoospores, but not for vegetative growth. It may be that sterols provide precursors for hormones that control the development of antheridia and oogonia; however, the closest we have to a structural determination is research that indicates both hormones to be neutral lipids with hydroxyl functional groups.

In the zygomycetes, typical members of which are the genera *Mucor*, *Rhizopus*, *Phycomyces*, and *Pilobolus*, the characteristic sexual structure is the zygophore, and sexual reproduction results in the formation of zygospores. In heterothallic species, when strains of different mating type [often called plus (+) and minus (−)] confront each other, the specialized branches called *zygophores* grow toward the opposing strain and make contact near their tips. They then swell at the point of contact, and these swellings develop into progametangia. A septum forms in each progametangium to separate the multinucleate gametangium from the supporting cell (suspensor).

The wall between the opposed gametangia is degraded and the fusion cell becomes the zygospore, which eventually develops a thickened wall. Karyogamy, or nuclear fusion, occurs in the zygospore, which is the only diploid phase in the life cycle. A zygospore germinates by forming a sporangiophore hypha with a single germ sporangium that contains many haploid, uninucleate spores. Many sporangia contain the products of just a single meiosis (amplified by postmeiotic mitoses), but the multinucleate gametangia can form several fusion nuclei, so the products of several meioses can appear in the spores. On the other hand, it may also be that some of the products of meiosis fail to be represented in the germ sporangia.

As is the case with the oomycetes, fusions between vegetative hyphae do not occur in zygomycetes; only zygophores fuse. As a result, zygophores must be highly differentiated hyphal branches. Development of zygophores is induced by trisporic acid in both mating types, and the hormone is active at concentrations as low as 10^{-8} M. Trisporic acid is formed by both plus and minus heterothallic strains in a confrontation, but its synthesis ceases if either partner is removed. The plus strain produces a precursor (methyl 4-dihydrotrisporate), which is converted to trisporic acid by the minus partner. The minus strain produces a different precursor (trisporol), which the plus strain converts to trisporic acid. Both precursors are volatile and their diffusion in the atmosphere is responsible both for induction of zygophores and for directing their growth toward one another in a mutual positive chemotropism.

Modern fungi *are* modern fungi and we do not want to fall into the trap of suggesting that what we have just described represent "primitive" or "early stages" in the development of fungal mechanisms for sexual reproduction. There are some lessons to be learned here, however, from modern

representatives of groups that diverged early in the evolution of fungi. First, both oomycetes and zygomycetes lack hyphal anastomoses, and generally only form hyphal septa to partition-off reproductive or injured structures. The independent syncytial nature of their mycelium creates the need (1) for specialized hyphal structures (like oogonia, antheridia, and zygophores) to be involved in sexual fusions, and (2) for mechanisms that induce and mutually attract these structures when potential mates encounter each other.

Both groups use chemical hormones. In the literature dealing with water molds these tend to be called *sex hormones* or *sex attractants*, but the word *pheromone* is increasingly applied and will be met in Section 2.6, and in later sections, in relation to mating processes in yeasts and filamentous fungi. A pheromone is a chemical emitted into the environment by an organism as a specific signal to another organism, usually of the same species. Pheromones, which are often effective at minute concentrations, play particularly important roles in the social behavior of animals, especially that of insects and mammals, being used to attract mates, mark trails, and promote social cohesion and coordination in colonies. This seems to be exactly what the sex hormones do in fungi, so the word has, not surprisingly, been adopted. It is now applied to a range of phenomena involved in the fungal mating process in which a molecule is released into the intra- or extracellular environment, and is then recognized by a specific pheromone receptor, which is a transmembrane protein. Through a G-protein (which uses GTP), the complex of pheromone + pheromone receptor then triggers a protein kinase signal transduction pathway that eventually specifically regulates gene activity (see Section 10.11).

The higher fungi (i.e., Ascomycotina and Basidiomycotina) have taken up and adapted the same strategy; for example, the trichogynes (specialized mating hyphae) of *N. crassa* protoperithecia grow toward hyphae of opposite mating type in response to a pheromone. There are some important differences in life style, however, between higher fungi and other eukaryotes that have great genetic significance. First, many fungal mycelia can tolerate (in fact, more than tolerate, can benefit from) the presence of several genetically distinct nuclei within their hyphae. This probably arises from the second important difference; namely, the fact that hyphal anastomoses occur very readily within the Ascomycotina and Basidiomycotina.

Hyphal anastomosis is the fusion between hyphae or hyphal branches, the process involves breakdown of two hyphal walls and union between two separate plasma membranes to bring the cytoplasms of the fusing hyphae into continuity with each other. Once they are in continuity, they can exchange nuclei and other organelles. The important point is that anastomosis is not limited to sexual reproduction; rather, hyphal fusions are essential to the efficient functioning of the mycelium of filamentous ascomycete and basidiomycete fungi because they convert the initially

radiating system of hyphae into the fully interconnected (and three-dimensional) network. Hyphal fusions are common within the individual mycelium as it matures. The interconnections they establish enable transport of nutrient and signaling molecules anywhere in the colony.

In unicellular fungi (i.e., the yeasts), fusion between the individual cells is called *conjugation*. This is very similar to the fertilization process that is the first step in the sexual cycle of animals and plants. Yeast cell conjugation also requires many of the processes required for fusion of filamentous vegetative hyphae: signaling by diffusible substances, directed growth, attachment of the two cell types to one another, production and targeting of hydrolytic enzymes to the attachment site, fusion of the plasma membrane, and restructuring of the cell wall to form a connecting bridge between the two cells. Database searches of the *Neurospora* genome for sequences involved in these processes in yeast have revealed a number of potential *N. crassa* homologues. Some of the machinery involved in yeast-cell fusion during mating may also be used for hyphal anastomosis in filamentous fungi.

Higher fungi are evidently fully equipped with the machinery necessary for hyphal tips to target other hyphae and to produce and externalize the enzymes (for cell wall degradation, for example) needed for anastomosis. This is part of normal mycelial development; it is not a specialization of the sexual cycle. It is complex machinery, though, because during normal mycelial growth vegetative hyphae usually avoid each other (known as *negative autotropism*). This behavior pattern promotes exploration and exploitation of the available substrate. Anastomosis requires that hyphae grow toward each other (called *positive autotropism*). It is not known how and why the usual avoidance reactions between hyphae become reversed, but it is evidently a change in the behavior of hyphal tips that occurs as the mycelium matures (and may depend on local population density).

Higher fungi, therefore, have a mechanism that promotes cell fusions to an extent that is never encountered in animals and plants, nor oomycetes and zygomycetes. Their problem is to regulate hyphal anastomosis so that its physiological and genetic advantages can be realized without hazard. And there are hazards. Hyphal anastomosis carries the risk of exposure to contamination with alien genetic information from defective or harmful cell organelles, viruses, or plasmids.

If genetically different hyphae are to interact (e.g., to take advantage of the sexual cycle), nuclear and cytoplasmic control requirements are very different. To maximize the advantage of sexual reproduction, the controls must ensure that the nuclei are genetically as *different* as possible. In contrast, safe operation of the cell requires that cytoplasms that are to mingle must be as *similar* as possible. These features are under the control of genetic systems that regulate the ability of hyphae to fuse, which are generally called *vegetative compatibility*. Genes called *mating-type factors*

subsequently regulate the ability of nuclei that have been brought together to undergo karyogamy and meiosis. It is important to recognize that these different aspects are recognizable phenotypes.

The phenotype of vegetative compatibility (also called vegetative, somatic, or heterokaryon incompatibility) is formation of a joint heterokaryotic mycelium, and the phenotype of compatible interaction between mating type factors is occurrence of sexual reproduction. Although we have to describe these systems separately because they are different phenotypes, and, further, that they must be described in terms of population biology, genetics, and molecular biology, remember that they are interdependent functions. Most of what we know about compatibility genetics has been obtained from laboratory studies with model organisms. In nature, however, vegetative compatibility is different from, but has a controlling influence over, mating-type function in terms of both population structure and genetic diversity.

2.2 Population Biology Aspects of Compatibility Systems

Vegetative compatibility is controlled by one to several nuclear genes that limit completion of hyphal anastomosis between colonies to those that belong to the same vegetative compatibility group (usually abbreviated to v-c group). Members of a v-c group possess the same vegetative compatibility alleles. The type of vegetative compatibility that is most usual in fungi is called *postfusion incompatibility*. Hyphal anastomosis is promiscuous in fungi, but compatibility of the cytoplasms determines whether cytoplasmic exchange will progress beyond the first few hyphal compartments involved in the initial interaction. If the colonies involved are not compatible, the cells immediately involved in anastomosis are killed (Fig. 2.1). This strategy prevents transfer of nuclei and other organelles between incompatible strains, but if the incompatibility reaction is slow, a virus or cytoplasmic plasmid may be communicated to adjacent undamaged cells before the incompatibility reaction kills the hyphal compartments where anastomosis occurred.

Vegetative compatibility prevents formation of heterokaryons except between strains that belong to the same v-c group. Hyphal fusions bring together nuclei of different origins into the same mycelium, and their ability to undergo karyogamy and meiosis to complete the sexual cycle is determined by the mating-type factors. Evidently, the best life cycle strategy for the fungi is to produce some offspring sexually for generating novel genotypes to adapt to the unpredictable and fluctuating environment, *and* produce many offspring asexually to rapidly colonize the favorable environment (once it is found) and so establish territorial control in a competitive world.

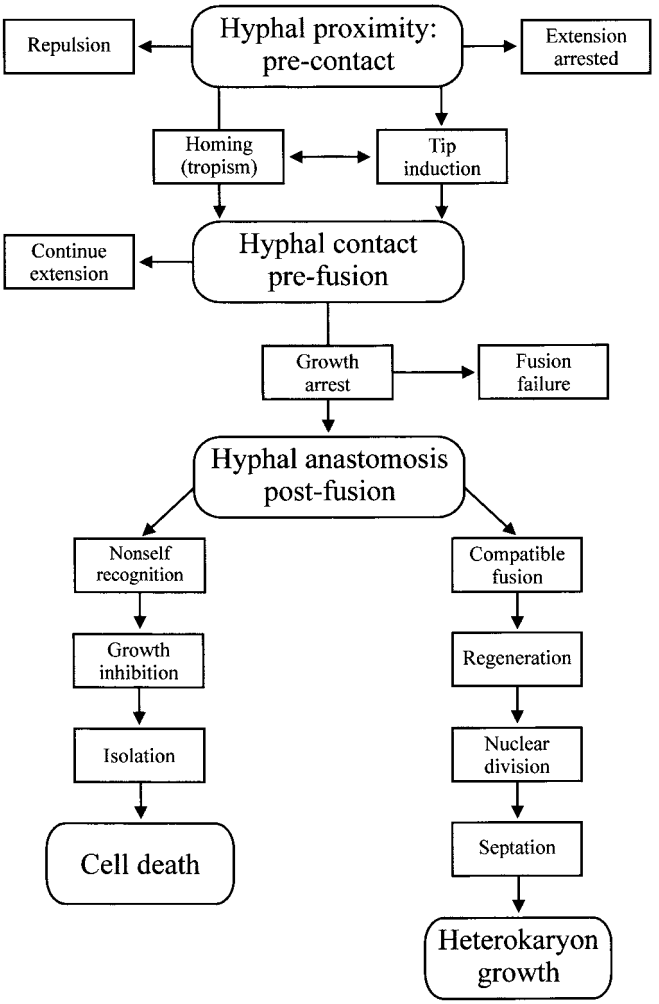


Fig. 2.1. Flow diagram illustrating the progress of hyphal interaction leading to operation of the vegetative compatibility systems. Recognition processes between hyphae take place at all three major steps: precontact hyphal proximity, prefusion hyphal contact, and postfusion self–nonself recognition. (Adapted and redrawn from Figure 1 in Glass, Jacobson & Shiu (2000), *Annual Review of Genetics* 34, p. 168.)

Mating (also called *breeding*) systems rely on nuclear genes that control mating between mycelia. Basic analysis of such systems depends on making experimental confrontations between mycelia and scoring whether or not the sexual stage is completed. It is, therefore, important to emphasize that such experiments test for the *phenotype* of sexual reproduction, and the pattern of its occurrence and its inheritance allow deductions about the control of sexual reproduction. A mycelium that possesses genes that prevent mating between mycelia that are genetically identical will be self-sterile; such a system is called *heterothallism* because it ensures that different mycelia must come together for a successful mating to occur.

Many heterothallic fungi, indeed all known heterothallic ascomycetes, have only two mating types specified by a single locus with different "alleles": *Neurospora crassa*, the brewer's yeast *Saccharomyces cerevisiae*, and the (basidiomycete) grass rust *Puccinia graminis* are examples. In such cases the mating type of a culture depends on which "allele" it has at the single mating-type locus (hence the alternative name of *unifactorial incompatibility*), where successful mating only takes place between cells or mycelia that have different "alleles" at the mating-type locus. Of course, the diploid nucleus that results is heterozygous for the mating-type factor, and meiosis produces equal numbers of progeny of each of the two mating types (hence the alternative name *bipolar heterothallism*). We put the word *allele* into quotes in the last few sentences because, although it is not evident from classical genetic analysis, one of the first things that molecular analysis revealed about the mating-type factors is that the different forms of the mating-type locus do not share the amount of DNA sequence homology you would expect of alleles. Their "alleles" can be very different indeed, differing in length in some cases by thousands of base pairs. For this reason they have been called *idiomorphs* rather than *alleles*. Idiomorphic structure (not allelism) is common to all fungal mating-type genes that are known.

With only two idiomorphs, the likelihood that two unrelated individuals will be able to mate (which is the outbreeding potential) is 50%. If there were n mating-type idiomorphs, however, then the outbreeding potential would be $[1/n \times (n - 1)] \times 100\%$; thus, the greater the number of mating-type idiomorphs, the greater the outbreeding potential. Many basidiomycetes have two unlinked mating-type factors (designated A and B); this is called a *bifactorial incompatibility system*. In this case, too, a compatible interaction is one between two mycelia with different idiomorphs, but this time both A and B mating-type factors must differ. As a result, the diploid nucleus that is formed will be heterozygous at the two mating-type loci, and meiosis will generate progeny spores of four different mating types (so *tetrapolar heterothallism* is the alternative name of this system).

Coprinus cinereus and *Schizophyllum commune* are the classic examples of this mating-type system. In both of these the wild population contains many different A and B idiomorphs, and the outbreeding potential is about

100%. The inbreeding potential of bifactorial incompatibility (the likelihood of being able to mate with a sibling) is 25% (because there are four different mating types among the progeny of a single meiosis), whereas it is 50% in unifactorial incompatibility where there are only two mating types among the progeny. A bifactorial system therefore tends to favor outbreeding. About 90% of higher fungi are heterothallic, and 40% of these are bipolar and 60% tetrapolar.

In homothallic (self-fertile) fungi, sexual reproduction can occur between genetically identical hyphae, but mating-type factors may still be involved. Primary homothallism occurs in species completely lacking heterothallism, but secondary homothallism occurs in species that have an underlying heterothallism that is bypassed. *Neurospora tetrasperma*, *Coprinus bisporus*, and *Agaricus bisporus* are good examples. In these cases, there are more postmeiotic nuclei than spores, so the spores become binucleate and heterozygous for mating-type factors. Spore germination gives rise to heterokaryotic mycelia that are, consequently, able to complete the sexual cycle alone (i.e., they act like homothallics). The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* exemplify a different process. Most strains are heterothallic with two mating types (see Section 2.6), but mating occurs in some strains between progeny of a single haploid ancestor (i.e., the culture appears to be "homothallic." The apparent homothallism results from a switch, in a few cells in the population, from one mating type to the other (see later) so that the (still heterothallic) clone comes to contain cells of different mating type.

It is the compatibility reactions (including, for the moment, both vegetative and mating-type compatibilities) that define in real life what constitutes the fungal individual. In yeasts each cell is clearly an individual, but a mycelial individual is not so obvious. Spores are individuals, and colonies developed from single spores must also be individuals, but are 10 spores from the same colony 10 different individuals, or just 10 bits separated from one individual? Then there are heterokaryons, which are mycelia that contain more than one nuclear type. Is a heterokaryon an individual, rather than a chimera or mosaic? These are important questions because in genetical terms a population consists of individuals that are able to interbreed. Individuals are important in evolution because selection operates on individuals. Thus, to understand fungal populations we have to know where the individual begins and ends.

Populations are important because the fundamental unit of biological classification, the species, is conventionally defined in terms of mating success and production of viable offspring. This is the biological species (or genospecies); in fungi, this is delimited by the experimenter carrying out mating tests between fungal isolates collected over a wide geographical area. This obviously applies to fungi that have a sexual process, but in Deuteromycotina, which lack sexual reproduction, the concept of a

genospecies is meaningless. In this latter case, individuals can be designated as a "taxospecies" on the basis of their similarity in many phenotypic characters. To achieve this molecular methods of analysis and computer-aided numerical analysis have become increasingly important. In most fungi, both types of analysis can be done and biological species do not always correspond to taxospecies because different selection pressures in different geographical areas may cause local populations to adapt into distinct taxospecies, which, despite obvious phenotypic differences, can interbreed when combined artificially (i.e., different taxospecies equals one genospecies). At the other extreme, mating barriers may arise, separating populations that are still sufficiently similar to remain in one taxospecies into different genospecies.

Most fungi have life cycles that consist of various phases, including haploid (n), dikaryotic (n + n) (or heterokaryotic with two or more types of nuclei in either a random or fixed ratio), and diploid (2n). Recessive alleles will not be expressed by heterozygous dikaryotic or heterokaryotic mycelia, and such mycelia may show hybrid vigor. The nuclei of a heterokaryon are open to selection and can segregate into homokaryons (or heterokaryons with altered nuclear ratios) if adverse conditions impose selection pressure. That is, an individual with a new phenotype might segregate from a heterokaryon in response to selection, the mechanism being selective segregation of some of its constituent nuclei into hyphal branches. This is a developmental process that allows the heterokaryon to adapt to its environment using an evolutionary mechanism; hyphal branches that contain a nuclear ratio best suited to the environment will be the ones to grow best. As a result, a heterokaryon is a very tolerant and adaptable mycelium. Heterokaryon interactions and interrelationships are also governed by their vegetative compatibility and mating-type systems. Most natural populations of fungi are composed of vegetatively incompatible, genetically and physiologically distinct individual heterokaryon mycelia that have been called genets.

Compatibility systems maintain the individuality of a mycelium and enable it to recognize unrelated mycelia of the same species with which it competes for territory and resources. In other words they provide an individual mycelium with a way of establishing whether hyphae it encounters belong to itself or not. When individuals do exchange nuclei it is the mating-type systems that then regulate sexual exchange between the mycelia by ensuring that compatible nuclei undergo karyogamy and enter meiosis. Separate genes for cytoplasmic compatibility and mating type can be identified in ascomycetes, but in basidiomycetes the genetic factors that have been identified as mating-type factors are complex multiple genes involved in determining self–nonself recognition, regulation of mycelial morphogenesis (the growth pattern of the dikaryon is different from that of the monokaryon), and regulation of karyogamy, meiosis, and multicellular development.

2.3 Compatibility and the Individualistic Mycelium

Fungal isolates from nature that are confronted with each other on artificial media usually show interactions implying self–nonself recognition. Such confrontations are the classic test for compatibility, which have been used for many purposes from population studies to allele complementation tests (see Section 4.1). They are set up by placing small pieces of the strains that are to be tested side by side on the surface of an agar medium. Tests should always be replicated; five replicates are best, but three will suffice. It is important that the confronted inocula are standardized as far as possible. They should be standardized in size and vigor by being cut to the same dimensions from a region about 2 mm behind the margin of a vigorous stock plate. The old medium should be removed from the inoculum before it is planted onto the fresh medium.

When the confrontations are incubated, leading hyphae may mingle and hyphal anastomoses occur between their branches. If the confronting strains are compatible, then the heterokaryon may proliferate so that the whole mycelium is heterokaryotic; this is what happens in *Neurospora crassa* and *Podospora anserina*. On the other hand, in species such as *Verticillium dahliae* and *Gibberella fujikuroi*, nuclei do not migrate between cells and heterokaryosis is limited to fusion cells. In this latter case, heterokaryons are continually formed by repeated fusion events within the mycelia. If the colonies involved are not compatible, the fusion cells are killed (Fig. 2.1). Cell death resulting from vegetative incompatibility involves plugging of the septal pores to compartmentalize dying hyphal segments, vacuolization of the cytoplasm, DNA fragmentation, organelle degradation, and shrinkage of the plasma membrane from the cell wall. It is an internalized cell death, different from necrotic cell death, with many features in common with programmed cell death (PCD, or apoptosis) in other multicellular eukaryotes.

Vegetative compatibility (= vegetative, somatic, or heterokaryon incompatibility) will prevent formation of a heterokaryon unless the strains belong to the same v-c group. Incompatibility between strains in a confrontation is caused by genetic differences between the two individuals at particular gene loci, which are called *het* (for *het*erokaryon) or *vic* (for *v*egetative *inc*ompatibility) loci, although once the major genes were identified several others that affected or otherwise modified their expression were also identified and given other descriptive names (Table 2.1). The *het* genes might define allelic (which are the most frequent type, and are also called *homogenic*) or nonallelic systems (*heterogenic*). In allelic systems, incompatibility is triggered by the expression of two incompatible alleles at the same locus. With nonallelic systems, incompatibility results from interaction between two genes at different loci. Only allelic systems have been found in *N. crassa* and *Aspergillus nidulans*, although both allelic and nonallelic systems occur in *P. anserina*. There are usually about 10 *het* loci, but the

Table 2.1. Genes involved in vegetative incompatibility that have been cloned and characterized.

Neurospora crassa	
Mat A-1	Mating-type transcription regulator, contains region similar to *Mat* α*1* of *Saccharomyces cerevisiae*
Mat a-1	Mating-type transcription regulator with an HMG box (characteristic of *H*igh *M*obility *G*roup proteins, which is a class of proteins distinct from histones that are found in chromatin and represent a subclass of the nonhistone proteins; the HMG proteins function in gene regulation and maintenance of chromosome structure)
het-c	Signal peptide (involved in endoplasmic reticulum targeting of secreted proteins) with glycine-rich repeats
het-6	Region of similarity to *tol* and *het-e* (of *P. anserina*)
un-24	Large subunit of type I ribonucleotide reductase
tol	Features a coiled-coil, leucine-rich repeat (a protein conformation found in extracellular matrix molecules), has regions similar to sequences in *het-e* (of *P. anserina*) and *het-6*
Podospora anserina	
het-c	Glycolipid transfer protein (glycolipids are involved in cell to cell interactions)
het-e	GTP-binding domain, region with similarity to *tol* and *het-6* of *N. crassa*
het-s	Prionlike protein (abnormally folded variant can infectively communicate its abnormal conformation to normal proteins that then form aggregates)
idi-2	Signal peptide, induced by *het-R*/*V* incompatibility
idi-1, idi-3	Signal peptide, induced by nonallelic incompatibility
mod-A	SH3-binding domain (*s*rc *h*omology domain 3; a protein domain of about 50 amino acid residues present in proteins involved in signal transduction, and also in a number of cytoskeletal proteins, generally involved in protein–protein interactions)
mod-D	α-subunit of G-protein with GTP binding (such proteins are involved in signal transduction in eukaryotic cells), modifier of *het-C*/*E* incompatibility
mod-E	Heat-shock protein (belongs to the Hsp90 family of 90 kDa polypeptides with ATPase activity that are essential for the viability of yeast cells and found in association with many regulatory proteins in eukaryotes, like steroid receptors and protein kinases), modifier of *het-R*/*V* incompatibility
pspA	Vacuolar serine proteinase, induced by nonallelic incompatibility

number varies from one species to another: there are at least 11 *het* loci in *N. crassa*, 9 in *P. anserina*, 8 in *A. nidulans*, and 7 in *Cryphonectria parasitica*. Two different alleles of a *het* gene are usually found in wild-type isolates, although *het* loci with more than two alleles have been found.

In *Aspergillus nidulans*, conidial color mutants have been induced in isolates from soil to study the extent of heterokaryon formation between pairs of strains. This approach works because the color of the conidium depends on expression of genes in the single haploid nucleus the spore contains; therefore, although the conidial head of *A. nidulans* is formed as spores bud off in chains from a multinucleate vesicle at the top of the conidiophore that could, in a heterokaryon, contain two kinds of nuclei, each chain of spores will be of one color. If the conidiophore is heterokaryotic, then coni-

dial heads of mixed color are produced; homokaryotic conidiophores have only one kind of nucleus in the vesicle and all their chains of conidia will be the same color. In one study, 18 different isolates were collected from around Birmingham, UK, and from Durham, which is about 300 km away. When confronted with each other, incompatible strains formed either none or >0.2% mixed conidial heads, whereas those that formed heterokaryons readily produced up to 5% mixed heads. In the former case the strains were considered to belong to different compatibility groups (v-c groups); in the latter, to the same group. Five v-c groups were identified among the 18 isolates, the largest of which included seven isolates that were collected from both localities.

Subsequent genetic analysis revealed that eight genes (called *hetA*, *hetB*, *hetC*, etc.) determine a total of six different heterokaryon compatibility groups in the homothallic *A. nidulans*. A difference at any of the *het* gene loci normally prevents heterokaryon formation and effectively defines the individual of *A. nidulans*. An important function of vegetative compatibility might be the protection of a mycelium from invasion by "diseases" following hyphal anastomosis. For example, an infectious cytoplasmically determined condition called "vegetative death" has been shown to be transmitted readily within a v-c group in *Aspergillus amstelodami*, but is only rarely transmitted between v-c groups.

A similar circumstance has been identified in *Ophiostoma ulmi*, the fungus that causes Dutch elm disease. Isolates of *O. ulmi* fall into mutually incompatible v-c groups, which differentiate races, subspecies, and even biological species. In compatible (called *c-type*) reactions between isolates there is no visible reaction or only a slight increase in density of the mycelium along the junction line between confronted cultures. A virus that causes a cytoplasmically transmitted disease of *O. ulmi* is transferable between strains that could establish c-type reactions, implying that hyphal anastomoses establish cytoplasmic continuity. Compatible reactions are rare (because most isolates tested are distinct individuals) and the most common reaction is an antagonistic one in which a broad mycelial barrage forms between the confronted cultures. This is the *w* or *wide* reaction. The disease virus is not transmitted between strains in a w reaction, which suggests that there is no cytoplasmic continuity. Narrow (n), "line" (l), and "line-gap" (lg) reactions are other antagonistic reactions that are sometimes observed. These reactions are different phenotypes of a vegetative compatibility system controlled by at least three genes that are functionally and genetically independent of the mating-type locus. In the wide reaction, hyphae penetrate from one colony into the space occupied by the other, which can lead to sexual recombination between the individuals defined by vegetative compatibility. Genetic exchange is usually unilateral: a strain may be an efficient female parent in crosses, but an inefficient male parent.

Neurospora crassa is similar to *Aspergillus* in that heterokaryon formation requires genetic identity at all *het* genes. One of these genes is the

mating-type locus of *N. crass*a. Although this is unusual, association between mating type and vegetative incompatibility is not restricted to *N. crass*a, but has been reported in *Ascobolus stercorarius*, *A. heterothallicus*, and *Sordaria brevicollis*.

In *N. crassa*, heterokaryons made between strains of opposite mating type grow slowly and have an irregular colony outline as compared with the rapid, uniform growth of heterokaryons between strains with the same mating type. The mating type gene of *N. crassa* evidently controls both sexual compatibility and heterokaryon compatibility, although the former requires that the mating types are different and the latter requires that the mating types are identical. It seems that nuclei of opposite mating type do not readily coexist in vegetative hyphae of *N. crassa*. Aggressive maintenance of individuality between mates is neither unusual nor difficult to understand in our own species; allegedly, men are from Mars, women from Venus. In *N. crassa*, the molecular basis of this mating aggression is that the MAT A-1 and MAT a-1 mating polypeptides (see Section 2.7) encode transcription regulators that specify different cell types in the sexual phase, but they are lethal when expressed together in a vegetative cell. The mating function of MAT a-1 depends on its DNA-binding ability, but this is not needed for the vegetative incompatibility function. Different functional domains of the polypeptide therefore serve these two different activities of the mating-type idiomorphs.

A functional *tol* gene is required for expression of this mating-type factor-mediated vegetative incompatibility in *N. crassa*. Incompatibility caused by the mating-type factor is suppressed by mutations in *tol* (these make the strains *tol*erant of each other), but *tol* mutations do not disrupt the sexual cycle. An active (i.e., nonmutant) *tol* allele is normally present in the heterothallic outbreeding species in which its mutant form was originally identified, but *N. tetrasperma*, which normally exists as a self-fertile (*MAT A* + *MAT a*) heterokaryon, has been shown to have an inactive *tol* allele.

Heterokaryons made between *N. crassa* strains of the same mating type (and the same *het* genotype) have nuclear ratios close to 1:1, full cytoplasmic continuity, and they produce up to 30% heterokaryotic conidia. In incompatible heterokaryon confrontations the pores in the septa of any cells that do fuse become blocked, and the cytoplasm becomes granular, then vacuolated, and finally dies. When such cytoplasm, or even a phosphate-buffer extract of it, was injected into a different strain, the same degenerative changes resulted. The activity of the extract was associated with its proteins, demonstrating that heterokaryon compatibility self–nonself recognition depends on the protein products of the *het* genes. The *un-24* gene of *N. crassa*, which is involved in *het-6* incompatibility, encodes the large subunit of ribonucleotide reductase, which is essential for DNA replication. Mutants in these *het* loci lose their nonself recognition ability and form vigorous heterokaryons with strains with which they were formerly incompatible.

Further insight into the mechanism of action of vegetative compatibility genes comes from work with *Podospora anserina*. When two incompatible colonies meet, hyphal fusion is followed by death of the fused cells and consequent absence of pigment, so a clear zone, called a *barrage*, forms between the colonies. The barrage is due to vegetative incompatibility, but the colonies might still be sexually compatible (controlled by one mating-type locus with idiomorphs called *plus* and *minus*). If they are compatible, a line of perithecia can be formed on each side of the barrage because lethality does not extend to fused trichogynes and spermatia, even though fused vegetative cells are killed. Analysis of the interactions of different races of *P. anserina* originally, revealed nine loci involved in barrage formation, several with multiple alleles. Incompatibility could result due to interaction between alleles of one gene, [e.g., allele t with t_1, u with u_1 (called *allelic* or *homogenic incompatibility*)] or to interaction between alleles of different genes [e.g., gene a_1 with gene b, gene c_1 with gene v (called *heterogenic* or *nonallelic incompatibility*)].

Heterogenic incompatibility reactions are often associated with fertility differences that, in some cases, are due to cytoplasmic factors transmitted through the hyphae. For example, interactions between the *R/r* and *V/v* genes indicate that the *V* gene product diffuses through the cytoplasm, whereas the *R* gene product is probably fixed to the plasma membrane. Observations like this give rise to the belief that vegetative incompatibility results from interaction of protein products of the incompatibility genes that trigger activity of proteins and enzymes during, and possibly causing, the senescence and death of the incompatible fusions (Table 2.1). Thus, the incompatibility genes are probably regulators of such enzymes. It has been argued that they may have arisen as mutations in genes normally concerned with housekeeping functions that ensure proper progress of the female (protoperithecial) part of the sexual cycle.

Several *het* loci of *P. anserina* have been characterized, including the allelic incompatibility locus, *het-s*, and the nonallelic *het* loci, *het-c* and *het-e* (Table 2.1). Note that *het-c* in *P. anserin*a has no relationship to the *het-c* of *N. crassa*. Just as in *N. crassa*, the *P. anserina het* loci evidently encode varied gene products. The *P. anserina het-c* locus encodes a glycolipid transfer protein; interestingly, disruption of *het-c* also impairs ascospore maturation. Although mutants of *het-s* exhibit no other phenotype than promiscuous vegetative compatibility, the *het-s* gene product behaves like a prion protein. A prion is a "proteinaceous infectious particle," a cellular protein that can assume an abnormal conformation that is infectious in the sense that it can convert the normal form of the protein into the abnormal (see Section 5.11). Hyphal anastomosis between *het-s* and the neutral *het-s** strain results in the cytoplasmic transmission and infectious propagation of the *het-s* phenotype.

Although the *het* loci encode very different gene products, three regions of similarity can be detected between predicted products of the *het-6* locus

and the *tol* locus of *N. crassa*, and the predicted product of the *het-e* locus of *P. anserina*. These regions might represent domains necessary for some aspect of vegetative incompatibility in which all three of these *het* loci are involved. Alleles of *het-c* that are found in *N. crassa* are present in other *Neurospora* species and related genera, which indicates that there was a common ancestor and conservation during evolution of this sequence. Despite this indication that there may be some underlying similarity in function, however, an *het* locus from one species does not necessarily confer vegetative incompatibility in a different species.

2.4 Nuclear Migration

The incompatible reaction evidently sets in train a large number of changes that result in the fusion cells being sealed off from the rest of the two interacting mycelia and killed. An equally dramatic change occurs following the compatible reaction in which hyphal fusion is followed, in many species, by migration of nuclei to give rise to a heterokaryon. Nuclei can migrate through pre-existing mycelium, either bidirectionally or unilaterally. The direction of migration depends on the genes and alleles involved in the compatibility reaction.

When confrontations are made between cultures of the two mating types of *Gelasinospora tetrasperma,* hyphae anastomose and nuclei migrate from one strain into the other. Perithecia develop along the junction line in one of the cultures because migration is regularly unilateral. Nuclear migration is fast: a typical migration rate of 4 mm per hour compares with a typical hyphal growth rate of only 0.7 mm per hour.

Nuclear migration has also been studied in some basidiomycetes. In the species that have been examined in most detail, *Coprinus cinereus* and *Schizophyllum commune*, basidiospores germinate to form homokaryotic mycelia with uninucleate cells, which is usually called a *monokaryon* or *homokaryon*. Two monokaryons will form hyphal anastomoses. If they are compatible, nuclei will migrate from one mycelium into the other. This establishes a new mycelium, called a *dikaryon*, which has regularly binucleate cells containing one nucleus of each parental type. The growth of dikaryotic hyphal tips requires that the two nuclei complete mitosis together (conjugate division), as well as a mechanism of nuclear migration and sorting that depends on a small backward-growing branch (called a *clamp connection* or *hook cell*) at each hyphal septum (Fig. 2.2). Other characteristic differences between the monokaryotic and dikaryotic mycelia of *C. cinereus* are that branches emerge from monokaryotic hyphae at a wide angle (40–90 degrees), but at an acute angle (10–45 degrees) from dikaryotic hyphae; monokaryons, but not the dikaryons, produce asexual arthrospores, called *oidia*, in droplets of fluid; and the aerial mycelium of monokaryons is generally less dense and fluffy than that of dikaryons.

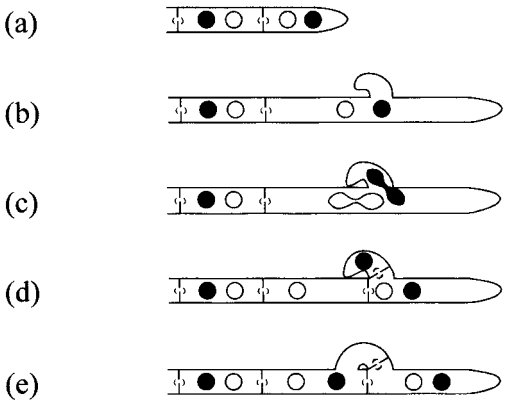
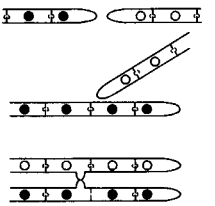


Fig. 2.2. Formation of a clamp connection. In dikaryotic hyphal cells (Fig. 2.2a) nuclei are paired, with one haploid nucleus of each mating type in each cell (here indicated by the closed and open circles). The apical cell extends, mitosis is initiated, and a clamp connection, or hook cell, emerges as a branch (Fig. 2.2b) that grows backward (away from the hyphal apex). Conjugate mitoses then occur (i.e., the two nuclei enter mitosis in synchrony), one daughter nucleus enters the clamp connection (Fig. 2.2c), and the other daughter nucleus of opposite mating type migrates away from the hyphal apex. Two septa are laid down (Fig. 2.2d), one traps a daughter nucleus in the clamp connection and the other divides the tip cell into two compartments at a location just behind the point of emergence of the clamp connection. Finally, the apex of the clamp connection anastomoses with its parent hypha (Fig. 2.2e) releasing its trapped nucleus into what is now the subterminal cell of the hypha.

Two multiallelic mating-type factors, called *A* and *B*, control sexual compatibility in *C. cinereus* and *S. commune*. To be compatible, the parental monokaryons must have different *A* and different *B* mating-type factors. The mating-type factors clearly determine that the initial self–nonself recognition that follows the first anastomoses of the encounter; they also regulate mycelial morphogenesis (presumably upregulating features that characterize the dikaryon and downregulating some monokaryotic features, like oidiation), but as it is only dikaryons that fruit in normal circumstances, the *A* and *B* mating-type factors also regulate fertility. This is why they are called *mating-type factors*. Dikaryon formation requires that both the *A* and *B* factors are different [which is taken to mean that both *A* and *B* are turned on (*A*-on, *B*-on, see Fig. 2.3)]. In a common-*A* heterokaryon the *A* factors are the same, but *B* factors are different (*A*-off, *B*-on) and nuclear migration can occur, but without conjugate divisions or clamp connection formation. When only *A* functions are turned on (*A*-on, *B*-off), in a common-*B* heterokaryon, conjugate divisions and clamp connection formation occur, but the clamp connections remain incomplete, and without clamp-cell fusion nuclear migration cannot take place (Fig. 2.3).

HYPHAL FUSION

Mating types do not influence hyphal fusion, which occurs irrespective of the mating type specificities of the interacting hyphae. Fusions might occur tip-to-tip, tip-to-side or side-to-side (by tip induction and outgrowth of peg-like branches).



A_1B_1 + A_2B_1	A_1B_1 + A_2B_2	A_1B_1 + A_1B_2
A-on, B-off	A-on, B-on	A-off, B-on

nuclear migration on

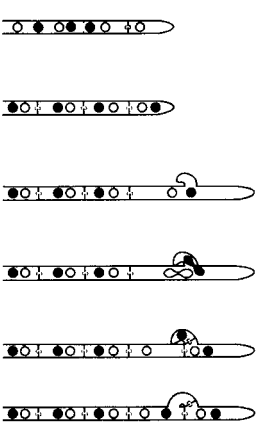
nuclear pairing on

clamp connection formation on

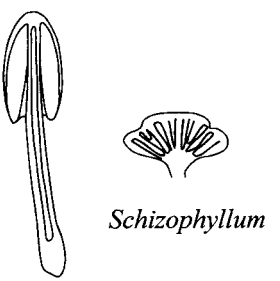
conjugate divisions on

clamp connection septation on

clamp connection tip fusion on



fruit body formation on



Coprinus

Schizophyllum

This "division of labor" between *A* and *B* factors is not universal; nuclear migration is regulated by the *B* factor only in *C. cinereus* and *S. commune*. Migration is regulated by both factors in *C. patouillardii*, but only *A* affects fertility. In *Ustilago* and *Tremella*, cell fusion is controlled by a locus with two idiomorphs, but the ability of fused cells to grow as dikaryons depends on a second locus with multiple idiomorphs (*Ustilago* mating types will be described in Section 2.8).

Nuclei migrate from a compatible mycelium and convert a monokaryon into a dikaryon. In experiments similar to those described earlier with *Gelasinospora tetrasperma*, it was found that nuclei of *Coprinus radiatus* (a close relative of *C. cinereus*) invade mycelia in such circumstances at a rate of 1.5 mm per hour, which is at least four times higher than the hyphal growth rate. In *C. congregatus*, a nuclear migration rate of 4 cm per hour (yes, *centimeters*!) has been reported, and migration rates in *Schizophyllum commune* range up to 2.7 mm per hour with a hyphal growth rate of only 0.22 mm per hour.

The invading nucleus undergoes regular division during nuclear migration. One of the daughter nuclei moves into the next cell, the intervening dolipore septa between adjacent cells of the monokaryon are broken down into simple pores through which the nuclei can be squeezed. Although some corresponding cytoplasmic movement has been observed in *Schizophyllum commune*, nuclear migration occurs more commonly without visible cytoplasmic flow. Nuclear migration is a highly active transport process involving microtubules in a manner analogous to the involvement of spindle fibers in the movement of chromosomes during division, which presumably accounts for its specificity. In most cases, a specific nuclear type is clearly being transported in a specific direction. In all this discussion of nuclear migration it is important to emphasize that *only* nuclei migrate; mitochondria are not exchanged between compatible mycelia. During migratory dikaryotization, anucleate cells and multinucleate cells are observed, so the dikaryotic state is not set up as soon as compatible cells fuse. Rather, ordered dikaryotic growth emerges after an interval of disorganized and irregular growth.

Fig. 2.3. Flowchart diagram of *A* and *B* mating-type factor activity in the basidiomycetes *Coprinus cinereus* and *Schizophyllum commune*. From top to bottom the flow chart depicts the events that take place when two haploid mycelia confront each other. Dikaryon formation requires that both *A* and *B* factors are different [which is taken to mean that both *A* and *B* functions are turned on (*A*-on, *B*-on)]. This is depicted in the central vertical line for a confrontation of $A_1B_1 + A_2B_2$, the vertical line connecting "function boxes," the phenotypes of that are indicated in the cartoons on the right-hand side. When only A functions are turned on (A-on, B-off), in a common-*B* heterokaryon ($A_1B_1 + A_2B_1$) conjugate divisions and clamp formation occur, but the clamp connections remain incomplete, and without clamp cell fusion nuclear migration cannot take place. In a common-*A* heterokaryon ($A_1B_1 + A_1B_2$) the *A* factors are the same, but *B* factors are different (*A*-off, *B*-on) and nuclear migration can occur, but without conjugate division or clamp connection formation.

2.5 Other Incompatibility Reactions

There are obviously numerous breeding systems and many genera contain species that exhibit a range of them. Although the tetrapolar system of *Coprinus cinereus* is one of the most extensively studied mating systems, other species of *Coprinus* are heterothallic bipolar, and some are homothallic. In the homothallic *Coprinus sterquilinus*, basidiospores germinate to form monokaryons, but after some growth they spontaneously convert to dikaryons with clamp connections.

Sistotrema brinkmannii exemplifies a different, but not uncommon, situation among the fungi. It is an aggregate species, isolates of which are difficult, even impossible, to distinguish on morphological grounds. The aggregate includes homothallic and heterothallic isolates, and the latter can be bipolar or tetrapolar. Homothallics fall into a single interfertile population, whereas bipolar and tetrapolar isolates fall into a number of intersterile groups. Hybrid progeny can be obtained from crosses between some homothallic and bipolar isolates, and gene segregation is quite regular. This pattern of behavior suggests that homothallism was derived from a bipolar heterothallic system by loss or inactivation of the mating-type factor; therefore, the incompatibility systems can evidently change as populations diverge in evolution.

The tone of the descriptions so far might imply that mycelial interactions take place only between homokaryons, but this is far from the truth. Study of the interactions *in vitro* is certainly easier with homokaryons, but when fungal mycelia are isolated from nature, the overwhelming majority proves to be heterokaryotic. Thus, in the real world the most normal interactions are those between heterokaryons. For example, a monokaryon of *Coprinus* can be dikaryotized by a dikaryon as well as by another monokaryon. In these "di-mon" (dikaryon × monokaryon) matings, if both dikaryon nuclei are compatible with the monokaryon (e.g., $[A_1B_1 + A_2B_2] \times A_3B_3$) either dikaryon nucleus (but generally only one of them) migrates into the monokaryon; if only one dikaryon nucleus is compatible with the monokaryon ($[A_1B_1 + A_2B_2] \times A_1B_1$), then the compatible nucleus will migrate into the monokaryon; and if neither dikaryon nucleus is compatible with the monokaryon ($[A_1B_1 + A_2B_2] \times A_1B_2$) there is often no migration. In about half the cases, however, where this confrontation was done in *Schizophyllum* it was found that the monokaryon became dikaryotized. In about half of those, both dikaryon nuclei migrated into the monokaryon, and in the others a mitotic recombination event in the dikaryon had produced a new nucleus that was compatible with the monokaryon. Common-*A* and common-*B* heterokaryons can also donate nuclei to monokaryons, as long as the monokaryon's nucleus has a different *B* factor (because *B*-functions are required for nucleus migration). Dikaryons and common-*B* heterokaryons of *Coprinus cinereus* can accept nuclei, but in *Schizophyllum commune* they only donate nuclei.

Diploid strains of *Coprinus cinereus* arise from common-*A* heterokaryons when rare nuclear fusions form a diploid nucleus that is packaged into an oidium. Thus, the common-*A* heterokaryon, $[A_1B_1 + A_1B_2]$ would give rise to a diploid $A_1A_1B_1B_2$. This diploid behaves like a monokaryon and could form dikaryons with monokaryons of mating-type A_2B_1 and A_2B_2. Dikaryosis requires that different factors are present in the two nuclei; that is, it is nonself idiomorphs that are compatible rather than self-idiomorphs being incompatible. Such diploids of *Coprinus* can only act as donors, not recipients, of migrating nuclei. Migration is sometimes unilateral in conventional monokaryon + monokaryon dikaryosis. The inability to accept nuclei is not related to the mating-type factor, but it is regulated by other genes.

Armillaria mellea has a tetrapolar mating system, but it is unusual in normally having one diploid nucleus in each cell of its rhizomorphs and in most cells of its fruit body stem and cap. In the gill tissues, however, there are multinucleate cells from which arise dikaryotic hyphae with paired nuclei and clamp connections. Young basidia are also binucleate, and seem to undergo a perfectly normal karyogamy and meiosis. Nuclei in the binucleate cells seem to be haploid, but how the haploids are derived from diploid nuclei in fruit body tissue is unknown. On the other hand, compatible matings between haploid spores give rise to the uninucleate diploid as the only stable mycelial growth, with the diploid condition arising as initial conjugate divisions form unusual patterns that result in uninucleate, diploid, cells. It is interesting that there is a recessive mutation of *Schizophyllum commune*, called *dik*, which converts dikaryotic hyphae into diploids when homozygous.

All of the emphasis in studies of incompatibility in *Coprinus cinereus* and *Schizophyllum commune* has centered on the mating-type factors. Indeed, deliberately inbred strains of *S. commune* have often been used to provide an isogenic background in which the gene of interest can be the only one that differs between test strains. A few closely related strains of *C. cinereus* have similarly been used for the bulk of the research in the hope of minimizing complications arising from the genetic background. Both of these strategies minimize or even abolish vegetative incompatibility, so it is not surprising that incompatibilities due to genes other than mating-type factors are rarely, if ever, encountered with these "model organisms." This does not mean that vegetative incompatibility does not occur in basidiomycetes.

Coprinus bisporus is bipolar in terms of mating type, but it shows additional incompatibilities due to interactions between other loci; this is similar to the heterogenic incompatibility of *Podospora* in which two monokaryons are incompatible if one has an allele C2 and the other allele D2. The combinations C1 + D1, C1 + D2, and C2 + D1 are compatible (as long as the mating-type factors are also a compatible combination). This heterogenic incompatibility was found to be common in isolates of the species collected

from nature. *C. bisporus* is secondarily homothallic and its dikaryons do not normally donate nuclei to each other. These features tend to preserve heterozygosis in the population, and the heterogenic incompatibility reinforces isolation between strains. *Heterobasidion annosus*, which is bipolar, has incompatibility due to five genes, each with two alleles called plus (+) and minus (–), superimposed on the *A* factor compatibility reaction. Compatibility requires that the confronting strains both have the plus allele of any one of these five genes (homozygosity for the minus allele does not confer compatibility). The system results in the natural population being comprised of several intersterile groups, genetically isolated from each other, and with the potential to show further evolutionary divergence.

In *Stereum hirsutum*, both the primary mycelium that emerges from a single basidiospore, and the secondary mycelium, which is formed by hyphal fusions between compatible primary mycelia, have multinucleate cells with multiple clamp connections at the septa. In this species (and other species of *Stereum*, *Phanerochaete*, and *Coniophora*) two different breeding strategies occur, called *outcrossing* and *non-outcrossing*. Some non-outcrossing reactions between nonsibling primary mycelia result in a macroscopic rejection response where the mycelia confront each other. This can be used to assign isolates to non-outcrossing interaction groups. Some non-outcrossing reactions, however, result in nonsibling mycelia intermingling without rejection. *Rejection* is taken to mean genetic difference between the isolates, but intermingling does not correspond to genetic identity.

For example, in interactions between three interaction groups of *S. sanguinolentum*, two isolates that rejected each other both intermingled with the third. Outcrossing interactions between nonsibling primary mycelial isolates produce a range of recognition responses that encompass compatibility, resulting in general or localized formation of a stable secondary mycelium. Rejection is signified by the formation of narrow demarcation zones of sparse, pigmented mycelium (with numerous vacuolated cells) between the colonies and the inhibition of extension of the colony margin of one or both of the mycelia, or the formation of flattened (called *appressed*) mycelium containing abnormal hyphae beyond the interaction interface. The pattern of formation of secondary mycelium indicates a multiallelic unifactorial (that is, bipolar) mating-type system on which the other phenotypes (controlled by several other genes) are superimposed.

We have described a wide range of incompatibility reactions in the last few sections. You will have noticed that what is true for one organism may not apply to another, even closely related, species. Do not get confused. Remember two things. First, fungi are very flexible organisms and those we find in nature today have each evolved a strategy to manage their interactions that best suits them in the environment in which they are competing. Second, in studying that strategy we need to use a standard approach that involves confronting one mycelium with another, and then studying what happens. We do this in the hope of recognizing general categories that

describe the manner in which the fungi interact, but in many cases our understanding is only partial, and may even be biased by the techniques chosen to make the observations. The crucially important point is that individuality *is* of prime importance to fungi and they have many ways of expressing their individuality.

2.6 Structure and Function of Mating-Type Factors: Mating-Type Factors in *Saccharomyces cerevisiae*

The life cycle of the yeast *Saccharomyces cerevisiae* features an alternation of a haploid phase with a true diploid phase. In this respect it differs from filamentous Ascomycotina in which the growth phase after anastomosis is a heterokaryon. There are two mating types: haploid yeast cells may be of mating type a or α. Karyogamy (nuclear fusion) follows the fusion of cells of the opposite mating type, and then the next daughter cell that is budded off contains a diploid nucleus. Most natural yeast populations are diploid because the haploid meiotic products mate while they are still close together immediately after meiosis. Diploid cells reproduce vegetatively by mitosis and budding until particular environmental conditions (e.g., deficiency in nitrogen and carbohydrate, but well aerated and with acetate or other carbon sources that favor the glyoxylate shunt) induce sporulation. When that happens, the entire cell becomes an ascus mother cell; meiosis occurs and haploid ascospores are produced. Ascospore germination reestablishes the haploid phase, which is itself maintained by mitosis and budding if the spores are separated from one another (i.e., by an experimenter in a laboratory, or by some disturbance in nature) to prevent immediate mating (Fig. 2.4).

Mating-type factors of yeast specify peptide hormones; these are called *pheromones* (the term originally applied to mate-attracting hormones of insects and mammals) and there are both pheromone α- and a-factors (Fig. 2.5), as well as corresponding receptors specific for each pheromone. Pheromones organize the mating process; they have no effect on cells of the same mating type or on diploids, but their binding to pheromone receptors on the surface of cells of opposite mating type (Fig. 2.4) act through GTP binding proteins to alter metabolism and: (1) cause recipient cells to produce an agglutinin that enables cells of opposite mating type to adhere; (2) stop growth in the G1 stage of the cell cycle; (3) and change wall structure to alter the shape of the cell into elongated projections. Fusion eventually occurs between the projections.

The mating process of *S. cerevisiae* is controlled by a complex genetic locus called *MAT* at which two linked genes are located (a1, a2 for mating type a and α1, α2 for mating type α). The *MAT*a locus encodes a1 and a2 polypeptides, the messengers for which are transcribed in opposite directions (Fig. 2.6), and *MAT*α encodes polypeptides α1 and α2. Heterozygos-

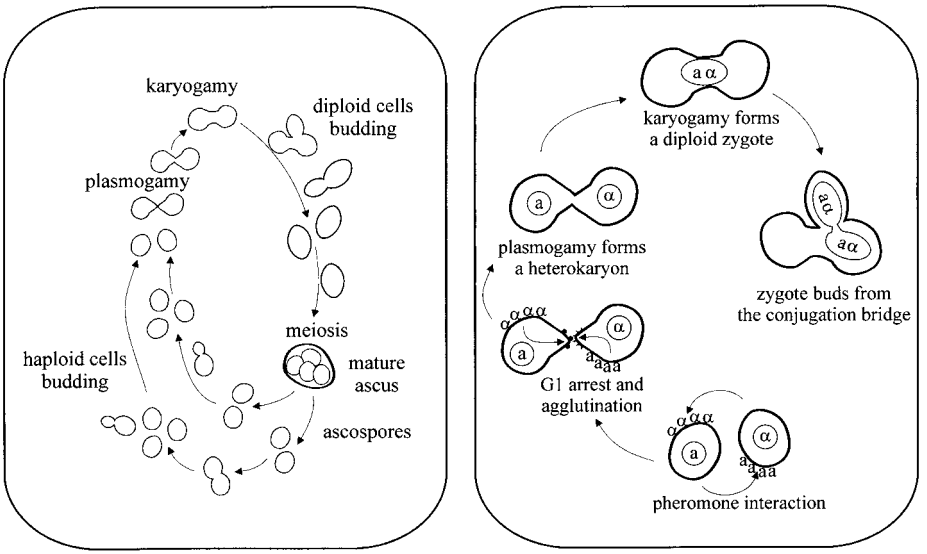


Fig. 2.4. Life cycle (left-hand panel) and mating process of the yeast *Saccharomyces cerevisiae*. Yeast can reproduce asexually by budding. Haploid cells of different mating types fuse to form dumbbell-shaped zygotes, which can themselves bud to establish a diploid clone. Well-nourished diploid cells, which are exposed to starvation conditions, enter meiosis, forming a four-spores ascus. Ascospores germinate by budding. In the laboratory, ascospores can be separated to form haploid clones, but in nature ascospores usually mate immediately, so the haploid phase is greatly reduced. The right-hand panel depicts pheromone interaction, agglutination, and the mating process in a little more detail.

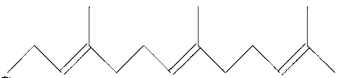
ity at *MAT* is a sign of diploidy and eligibility to sporulate; even partial diploids carrying *MAT*a/*MAT*α will attempt to sporulate. In haploid cells, the α2 polypeptide represses transcription of a-factor in α-cells, whereas a1 represses α-specific genes in a-cells. The α1 protein activates transcription of genes coding for α-pheromone and the surface receptor for the a-factor. In a–α diploids, interaction occurs between a1 and α2 polypeptides to form a heterodimer, which represses genes specific for the haploid phases, including a gene called *RME*1, which itself suppresses meiosis and sporulation (Fig. 2.6).

Saccharomyces cerevisiae is heterothallic, but a clone of haploid cells of the same mating type frequently sporulates, and there will be equal numbers of a and α cells among the progeny. This results from mating-type switching controlled by the gene *HO* (HOmothallic) that exists in two allelic forms (dominant *HO* and recessive *ho*), and encodes an endonuclease. There are silent storage loci for each mating type, called *HML* and *HMR*, on either side of the *MAT* locus, and on the same chromosome. The *HO–ho* endonuclease creates a double-strand break at the *MAT* locus that

α-factor

NH_2-Trp-His-Trp-Leu-Gln-Leu-Lys-Pro-Gly-Gln-Pro-Met-Tyr-COOH

a-factor



NH_2-Tyr-Ile-Ile-Lys-Gly-Val-Phe-Trp-Asp-Pro-Ala-Cys-$COOCH_3$

Fig. 2.5. Simplified chemical structures of yeast pheromones.

initiates switching of information, by a homologous recombination event between the two parts of the same chromosome, at the *MAT* locus with that at either *HML* or *HMR* (Fig. 2.7).

Because yeasts can live in very small habitats, like flower nectaries and surfaces of individual fruits, rare mating-type switching will give isolated populations the opportunity to undergo sexual reproduction; presumably, this is its selective advantage. Mating-type switching occurs about once in

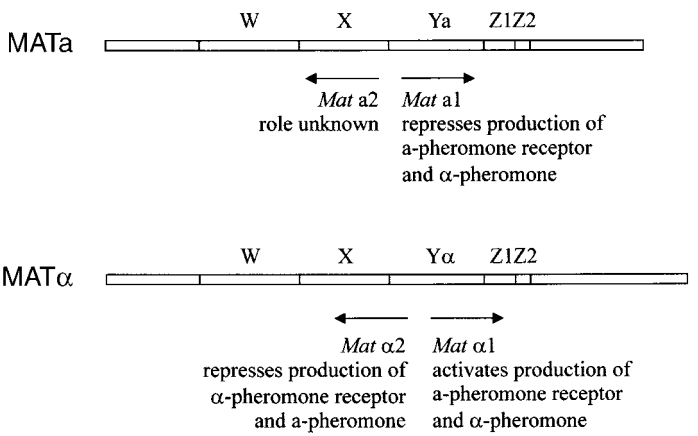


Fig. 2.6. Functional domains in mating-type factors of *Saccharomyces cerevisiae*. Region Y is the location of the mating-type idiomorphs, which have very little homology with each other. Ya is 642 bp long, Yα is 747 bp long. Regions W, X, and Z1 and Z2 are homologous terminal regions. The arrows indicate direction of transcription, and the legends beneath the arrows indicate functions of the gene products. In *S. cerevisiae* of mating type a, a general transcription activator is responsible for production of a-pheromone and the membrane-bound α-pheromone receptor. In a–α diploids, the MATa1/MATα2 heterodimer protein activates meiotic and sporulation functions, and represses haploid functions (turning off α-specific functions by repressing MATα1; a-specific functions are repressed by MATα2 alone).

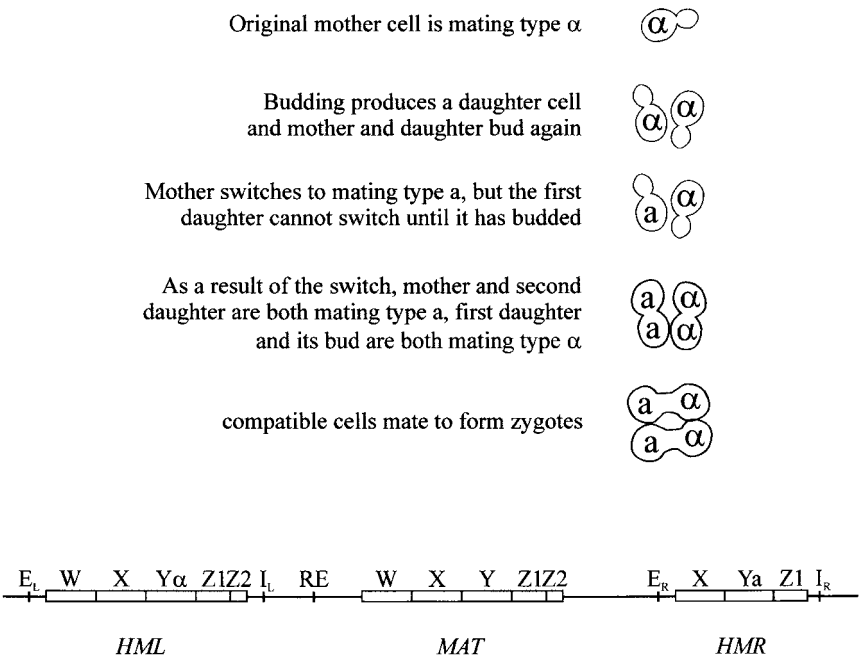


Fig. 2.7. (Top) pattern of mating-type switching in *Saccharomyces cerevisiae* showing the consequences of a mating-type switch in one mother cell. (Bottom) the three loci involved in mating-type switching, *HML*, *MAT*, and *HMR*, are located on the same chromosome (*not* drawn to scale). *HML* is about 180 kb from *MAT*, and *HMR* about 120 kb from *MAT*; the centromere is located between RE and the *MAT* locus. A double-strand break at the *MAT* locus, caused by the HO endonuclease, initiates a recombination event that replaces the Y region of the *MAT* locus with Y sequences from one of the storage loci. *HML* and *HMR* contain complete copies of the mating-type genes, but they are not expressed because they have a repressed chromatin structure imposed by the E and I silencer sequences. *HML* shares more of the *MAT* sequences (W, X, Z1 and Z2) than does *HMR*. RE is a recombination enhancer that controls preferential recombination between *MAT*a and *HML*, or between *MAT*α and *HMR*. (Figs. 2.3–2.7 adapted and redrawn from Chiu & Moore (1999), Sexual development in higher fungi. Chapter 8 in *Molecular Fungal Biology* (R. P. Oliver & M. Schweizer, eds.), pp. 231–271, Cambridge University Press.)

10^5 divisions in cultures carrying allele *ho*, whereas the switch occurs at every cell division in strains carrying *HO*. There is an asymmetry in the cell division, however, in that a new daughter bud is not able to switch mating types until it has itself budded. This is achieved in *S. cerevisiae* by actively transporting into the budding daughter cell the mRNA of a gene called *Ash1*, which encodes an inhibitor of the *HO*-endonuclease. As a result, immediately after each division only the mother cell is switchable, which means that even if there is only one cell to start with, a single division cycle will produce two cells of opposite mating type. If you think that's a clever arrangement, the switch to *opposite* mating type is assured because a

recombination enhancer controls recombination in the arm of chromosome III on which all these genes are located. This control region ensures that in *MATa* cells the resident *MATa* locus recombines with *HML*, which contains a silent *MAT*α locus, whereas in *MAT*α cells the resident locus recombines with *HMR*, which contains a silent *MATa* locus. Now, *that's* clever!

Mating-type switching also occurs in the distantly related fission yeast *Schizosaccharomyces pombe*, but this organism uses the asymmetry of DNA replication to establish an asymmetrical mating-type switching pattern. When *S. pombe* divides, the two daughter cells exhibit different developmental programs: one is mating-type switchable, the other is unswitchable. Genetic experiments show that the expressed (*mat1*) mating-type locus in switchable cells has an imprint that marks it as a candidate for the intrachromosomal recombination event that makes the mating-type switch. The imprint is a modification in one strand of the DNA, possibly a "nick" (i.e., a broken phosphodiester bond) or an RNA primer left from the DNA synthesis during the previous mitotic division. During DNA replication the strand-specific imprint is made at the *mat1* locus only during lagging-strand synthesis, so only one of the sister chromatids will carry the imprint. The cell that inherits the imprinted chromosome becomes a switchable cell, whereas its sister remains unswitchable. When the imprinted chromosome is replicated, the DNA replication complex runs into the imprinted modification in the DNA, the replication fork stalls, and the result is a transient double-strand break that initiates the recombination required for mating-type switching.

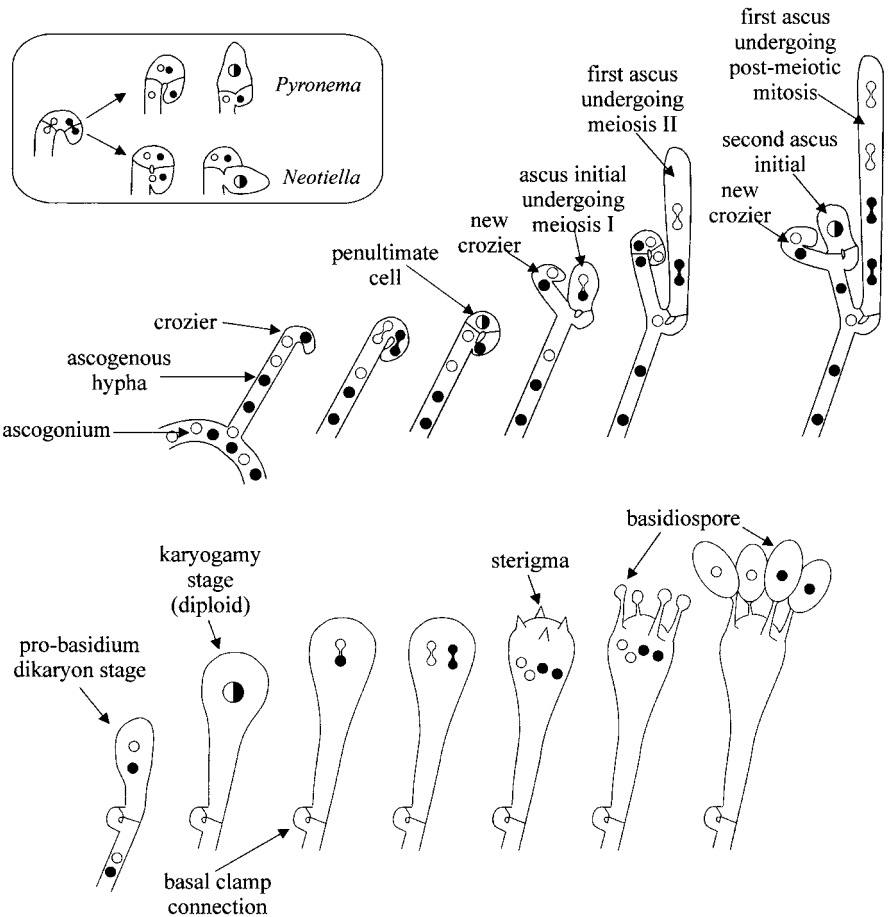
Mating types in filamentous fungi tend to be far more stable, although unidirectional switching of mating type has been reported in some filamentous ascomycetes, although with no molecular details yet. Switching oddly enough does not occur in any of the best-studied organisms like *Neurospora*, *Aspergillus*, or *Podospora*, but has been claimed in *Chromocrea spinulosa*, *Sclerotinia trifoliorum*, *Glomerella cingulata*, and *Ophiostoma ulmi*.

2.7 Structure and Function of Mating-Type Factors: Mating-Type Factors of *Neurospora*

Species of *Neurospora* exhibit four different mating strategies: (1) bipolar heterothallism with mating types *A* and *a* (in *N. crassa*, *N. sitophila*, *N. intermedia*, and *N. discreta*), but mating-type genes are present in a single copy per genome (unlike *Saccharomyces cerevisiae*); (2) secondary homothallism (in *N. tetrasperma*) through the production of asci containing four ascospores each containing compatible nuclei; (3) primary homothallism in which each haploid genome carries genetic information of both mating types (*N. terricola*, *N. pannonica*); and (4) primary homothallism, but in which genetic information for only one mating type can be detected (for

example, *N. africana* possesses only an *A* idiomorph that shows 88% homology with the *A* idiomorph of *N. crassa*).

Species that show primary homothallism form linear eight-spored asci (octads) in which all progeny are self-fertile. In the bipolar heterothallic species, strains of both mating types develop female structures (i.e., protoperithecia and their receptive hyphae, the trichogynes) under nitrogen starvation, as well as asexual spores (macroconidia or microconidia). Because the spores can serve as the male in a sexual cross these mycelia are hermaphrodites. Migration of (the still haploid) nuclei into the *N. crassa* ascogonium (see Fig. 2.8) depends on mating-type gene function. After the arrival in the ascogonium, a series of mitotic divisions occur. There must be some mechanism to sort nuclei (i.e., to ensure that meiosis only involves one *a* and one *A* nucleus) because mating type *always* segregates 1:1 in the progeny. It is likely that transient dikaryosis in the crozier involves a nuclear



recognition mechanism; however, crozier abortion occurs even in normal ascosporogenesis, so an alternative process might be that a nonselective nuclear migration results in abortion of croziers that do not contain one *a* and one *A* nucleus.

Mating is followed by the formation of perithecia (Fig. 1.1), within which as many as 200 asci are formed. Each ascus contains the products of a single meiosis. In many filamentous ascomycetes, a postmeiotic mitotic event, before ascospore formation, results in each ascus containing an octad comprised of four pairs of sister ascospores.

The best-characterized mating-type loci from filamentous ascomycetes are those of *Neurospora crassa*, *Podospora anserina*, and *Cochliobolus heterostrophus*. *N. crassa* and *P. anserina* mating-type loci contain essentially the same genes, but *C. heterostrophus* has a simpler arrangement. Mating-type genes of *Magnaporthe grisea* are probably similar to those of *N. crassa*.

The first mating-type genes to be cloned and sequenced in filamentous fungi were the *A* and *a* loci of *Neurospora crassa* (Fig. 2.9). In all the known *Neurospora A* and *a* idiomorphs the flanking regions are conserved, and the region on the centromere side contains species-specific and/or mating-type–specific DNA sequences. Immediately adjacent to these segments are

Fig. 2.8. Meiosis and sporulation in Ascomycota and Basidiomycota. The major panel of diagrams at the top shows ascus formation. Hyphal fusion or similar mating between male and female structures results in nuclei moving from the male into the female to form an ascogonium in which male and female nuclei may pair but do not fuse (transient dikaryon stage). Ascogenous hyphae grow from the ascogonium. Most cells in these hyphae are dikaryotic, containing one maternal and one paternal nucleus, with the pairs of nuclei undergoing conjugate divisions as the hypha extends. In typical development, the ascogenous hypha bends over to form a crozier. The two nuclei in the hooked cell undergo conjugate mitosis after which two septa are formed, creating three cells. The cell at the bend of the crozier is binucleate, but the other two cells are uninucleate. In general, the binucleate cell becomes the ascus mother cell, in which karyogamy takes place. In the young ascus meiosis results in four haploid daughter nuclei, each of which divides by mitosis to form the eight ascospore nuclei (the octad). The boxed inset shows that karyogamy may occur in the penultimate cell of the crozier (*Pyronema*-type) or the terminal and stalk cell nuclei might fuse (*Neotiella*-type). The panel of diagrams at the bottom shows basidium formation in a "classic" mushroom-fungus. The basidium arises as the terminal cell of a dikaryotic hyphal branch that inflates and undergoes karyogamy and meiosis. At the conclusion of the meiotic division four outgrowths (sterigmata) emerge from the basidial apex and inflation of each sterigma tip produces the basidiospore (which is an exospore, produced outside the meiocyte in contrast to the endospores of ascomycetes). Nuclei then migrate from the basidium into the newly formed basidiospores. Mitosis may take place within the basidiospores before they are discharged. Comparison of these diagrams indicates how tempting it is to suggest some evolutionary relationship between crozier formation and the early stages of basidium and clamp connection formation.

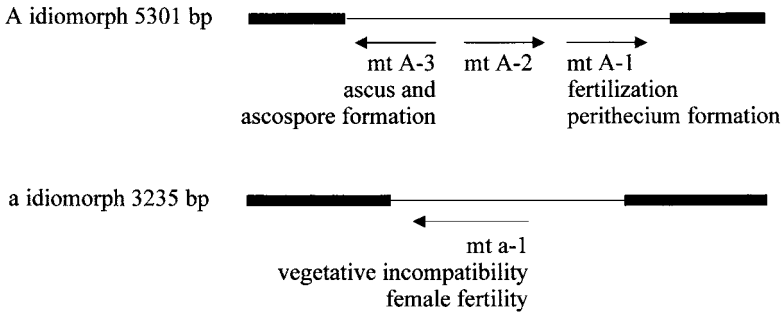


Fig. 2.9. Functional regions of mating-type factors of *Neurospora crassa*. The arrows indicate direction of transcription; the legends beneath the arrows indicate functions of the gene products. The black bars represent the conserved DNA sequences either side of the idiomorphs, which are shown as lines. These diagrams are oriented so that the centromere is on the left; consequently, the centromere-distal sequence is on the right.

regions that are very different between species. The idiomorphs themselves are next to these species-variable regions. These are highly conserved between species, but are completely dissimilar between the two mating types within the species. These are then followed by a "mating-type common region" of 57–69 bp, which separates an idiomorph from its nearby variable region and is very similar between species and between the two mating types.

The *A* idiomorph is 5301 bp in length and gives rise to at least three transcripts (MAT *A*-1, *A*-2, and *A*-3), where the first two are transcribed in the same direction (Fig. 2.9). The 85 amino acids at the N-terminal region of the mating-type *A* product are the minimum required for expression of female fertility. The region from position 1 to 111 determines the vegetative incompatibility activity of the mating-type locus, and amino acids from position 1 to 227 are required for male-mating activity. Mating-type–specific mRNA is expressed constitutively in vegetative cultures, and continues to be expressed after mating both before and after fertilization. Transcript MAT *A*-1 is very similar to *MAT*α1 of *Saccharomyces cerevisiae* and is essential for fertilization and fruiting body formation.

The other two transcripts, MAT *A*-2 and MAT *A*-3, increase fertility and are essential in events after fertilization, including ascus and ascospore formation, but are not essential for sexual development, which is controlled by MAT *A*-1. The MAT *A*-3 transcript has DNA binding ability and might function as a transcription-regulating factor. The function of the MAT *A*-2 polypeptide is not known. Its sequence contains motifs that occur in transcription activator proteins, but there are no obvious relatives in sequence databases other than the *Podospora* mating-type homologue, SMR1.

The *a* idiomorph of *N. crassa* is 3235 base pairs long and gives rise to a single transcript (called MAT *a*-1) that encodes one polypeptide of 288

amino acids with DNA-binding activity. MAT *a*-1 is the only gene essential to mating in the a idiomorph, and has a role in the dikaryon stage. Amino acids 216–220 of the MAT *a*-1 polypeptide act in vegetative incompatibility, whereas the region with DNA-binding activity is responsible for the mating function, which implies that vegetative incompatibility and mating work through different mechanisms. The DNA sequences bound by MAT *a*-1 center on CAAAG sequences, which is similar to "high mobility group" (HMG) proteins that bind in the minor groove of the DNA helix and introduce a bend in the DNA molecule. DNA-binding targets differ in different developmental stages, and the specificity may result from MAT *a*-1 interacting with unidentified protein factors. Mutations in either MAT *a*-1 or MAT *A*-1 cause mating defects.

N. crassa trichogynes respond to pheromones by orienting toward the pheromone source in a mating-type–specific way. Mating-type mutants do not orient their growth toward pheromones. DNA sequences of fungal pheromone receptors predict a product with seven transmembrane segments with the ability to interact with a heterotrimeric G-protein linked to a protein kinase cascade. Transcription of pheromone receptor genes is regulated by the mating-type factors (in basidiomycetes, pheromone receptors are products of a mating-type locus, see later). A G-protein encoding gene with mating function, called *gna-1*, has been characterized in *N. crassa*. Female fertility is lost if *gna-1* is disrupted. Female infertility also results if the *gna-1* homologue in *Cryphonectria parasitica*, which is called *cpg-1*, is disrupted. This suggests that this G-protein is a component of a female pheromone response pathway that has been conserved in these filamentous ascomycetes.

The *mat*+ mating-type idiomorph of *P. anserina* contains a single gene (called *FPR*1), which is homologous to *Neurospora* MAT *a*-1. The *mat*– idiomorph encodes three genes (*FMR*1, *SMR*1, *SMR*2) similar to MAT *A*-1, *A*-2, and *A*-3 of *Neurospora.* Functions of the *P. anserina* genes are very similar to those of their counterparts in *N. crassa*. The mating-type idiomorphs of *Cochliobolus heterostrophus* each contain only one gene. One of the idiomorphs contains a sequence homologous to the Mat-α1 of *S. cerevisiae*, and the other contains a sequence homologous to the *N. crassa* MAT *a*-1 gene.

2.8 Structure and Function of Mating-Type Factors: Mating-Type Factors in *Ustilago maydis*

Ustilago maydis causes the smut disease of maize. It has a tetrapolar mating system comprising one multiallelic mating type factor and one with only two alleles. *Ustilago* produces unicellular, haploid sporidia that grow vegetatively by budding like a yeast phase; these can be cultured on synthetic media and are nonpathogenic for the host plant (Fig. 2.10). Conjugation

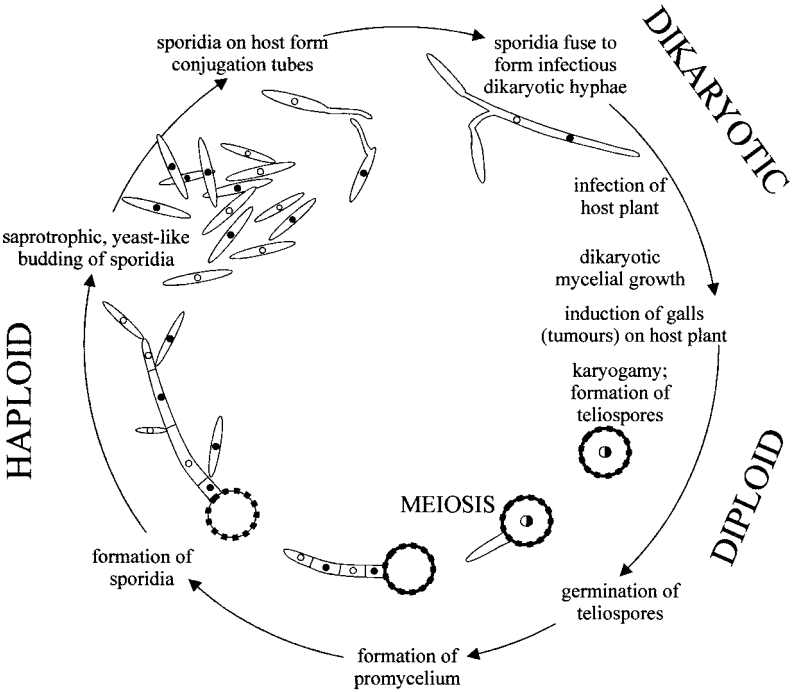


Fig. 2.10. Diagram of the life cycle of *Ustilago maydis*.

tubes are formed when sporidia of opposite mating type are mixed, and fusion of these produces the dikaryon, which then grows as a filamentous fungus. The dikaryon is the pathogenic stage. Fusion of sporidia is controlled by the biallelic "*a*" mating-type locus, the heterozygous *a*1/*a*2 genotype that is required for conjugation and the transition between the yeast and filamentous forms. The multiallelic "*b*" locus stops diploid cells fusing, determines the true hyphal growth form, and pathogenicity.

The *a*1 idiomorph consists of 4.5 kb of DNA and the *a*2 idiomorph is 8 kb. Two genes have been identified in these regions: *mfa*1 (in *a*1) and *mfa*2 (in *a*2) code for pheromones, and *pra*1/*pra*2 encode pheromone receptors. The pheromones diffuse away from their producer cells and induce conjugation tubes after binding to pheromone receptors on cells of opposite mating type. Pheromone signaling is also necessary for the maintenance of the filamentous dikaryon after cell fusion. Pheromone induces all of the mating-type genes to levels 10–50 times higher than their basal level. The upstream control element responsible for this pheromone stimulation, which is called the pheromone response element (PRE), has the sequence ACAAAGGG. It is on the same DNA molecule as are the pheromone genes, so it is called a "*cis*-acting element," by analogy with chemical terminology that describes two substituents on the same side of an axis of sym-

metry in a molecule as the *cis*-configuration, and the alternative of two substituents on opposite sides of an axis as the *trans*-configuration.

The PRE sequence is similar to the consensus sequence recognized by HMG polypeptides, including MAT *a*-1 of *N. crassa* (see Section 2.7). A gene called *pfr1* (*p*heromone *r*esponse *f*actor), the product of which binds to the PRE sequences found in both *a*1 and *a*2 idiomorphs, encodes the controlling transcription factor. The downstream pheromone response pathway includes at least one MAP kinase encoded by a gene called *fuz7*, which is homologous to the "archetypal" MAP kinase gene of *Saccharomyces cerevisiae*, *STE7* (a MAP kinase is a "mitogen activated protein kinase," where a mitogen is any agent that induces mitosis; see Section 10.11). Disruption of *fuz7* results in phenotypes that show that *fuz7* is involved in *a*-dependent mating events like conjugation tube formation, conjugation, and establishment and maintenance of filamentous growth. Other components of the pheromone response pathway that have been found are four genes encoding G-proteins (*gpa1* to *4*). There are indications that *fuz7* and *gpa3* do not belong to the same pathway, so there may be several pheromone responses in *U. maydis*, either in parallel or in series.

The *b* mating-type factor contains two genes that are transcribed in opposite directions (Fig. 2.11): *bE* and *bW* (= East and West) with coding sequences equivalent to polypeptides of 473 and 629 amino acids, respectively. The amino terminal end of the coding sequence is highly variable, whereas the carboxy-terminal end is conserved in different *b* idiomorphs. The *bE* and *bW* proteins are, respectively, the HD1 and HD2 homeodomain proteins that form a dimer that is a transcription activator of genes required for the sexual cycle and/or repressor of haploid-specific genes (interaction of the equivalent *Coprinus* homeodomain proteins is illustrated later in Fig. 2.14). Dimers comprised of *bE* and *bW* from the same idiomorph are inactive; the heterodimer functions properly only when the proteins come from different idiomorphs. The proteins encoded by these genes contain sequences homologous with DNA-binding homeodomain regions of known transcription-regulating factors, which is why they are called *HD1* and *HD2*. These may encode transcription factor themselves. It is significant that the homeodomain is an extended helix-turn-helix DNA-binding motif (see Section 10.3), which is encoded by a conserved DNA sequence of about 180 bp called the *homeobox*. This sequence is particularly associated with the transcriptional regulators of homeotic, or *Hox*, genes that were found originally in the fruit fly *Drosophila*, and which are involved in orchestrating development in higher eukaryotes. Mutations in animal *Hox* genes convert one body part into another. *Drosophila* has two *Hox* clusters, but vertebrates have four clusters of 9–11 genes each on different chromosomes. Vertebrate *Hox* genes are expressed in different patterns and at specific embryological stages. There is a compelling comparability between these developmental regulators and the *HD1*/*HD2* genes of the fungal mating-type factors in *Schizophyllum* and *Coprinus*, as well as *Ustilago*.

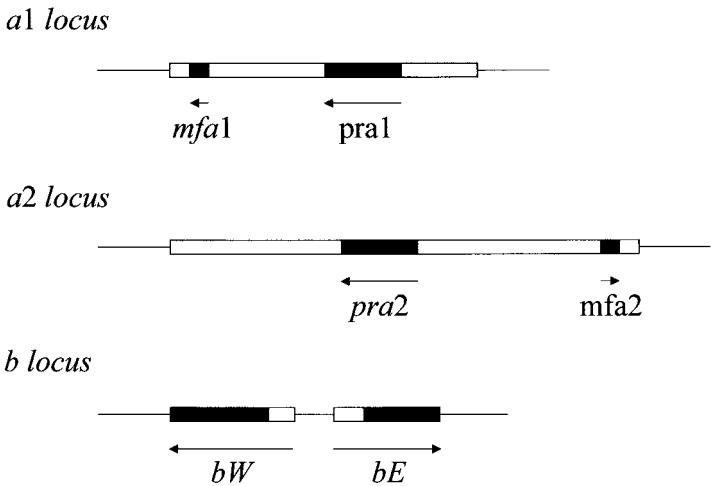


Fig. 2.11. Schematic representations of the structures of the *a* and *b* mating-type loci of *Ustilago maydis*. Idiomorphs of the *a* locus consist of mating-type specific (i.e., variable) DNA sequences (4500 base pairs in *a*1; 8000 base pairs in *a*2), here shown as open boxes, within which are the genes for mating (*mfa* and *pra*). The *b* locus has two reading frames, *b*W and *b*E, that produce polypeptides containing domains of more than 90% sequence identity (shown as black boxes) and variable domains (open boxes), which show 60–90% identity. Arrows indicate the direction of transcription. The PRE sequence mentioned in the text is a very short control site upstream of the pheromone genes, *mfa*1 and *mfa*2.

2.9 Structure and Function of Mating-Type Factors: Mating-Type Factors in *Coprinus cinereus* and *Schizophyllum commune*

C. cinereus and *S. commune* exhibit tetrapolar heterothallism that is determined by two mating-type factors, called *A* and *B*. The natural population contains many different mating types. In crosses these behave like multiple alleles of the two mating-type loci. Molecular analysis has revealed that each mating-type locus is a very complex region that contains several or even many genes (which is why we refer to mating-type *factors*). The genes at *A* encode transcription factor homeodomain proteins; genes at *B* encode lipopeptide pheromones and pheromone receptors. Mating-type factors are located on different chromosomes, and even conventional genetic analysis has demonstrated internal structure, identifying subloci that are called $A\alpha$, $A\beta$, $B\alpha$, and $B\beta$. These subloci are relatively far apart, in terms of linkage distance, in *S. commune*, but they are much closer in *C. cinereus* (Fig. 2.12). The α and β subloci are functionally redundant in the sense that a difference need exist at only one of them for compatibility. Nine versions of $A\alpha$ and 32 of $A\beta$ result in 288 different *A* mating-type specificities in *S.*

commune. There are an estimated 160 *A* mating-type specificities in *C. cinereus* in the natural population.

A compatible mating in both fungi, characterized by clamp connections and conjugate nuclear divisions in the mated hyphae, requires that both *A* and *B* are different. In the belief that in this state the mating-type factors are fully active, it is called *A*-on, *B*-on (Fig. 2.3). Cytological observations indicate that mating-type factor *A* controls nuclear pairing, clamp cell formation, and the synchronized (conjugate) mitosis of nuclei, whereas mating-type locus *B* controls nuclear migration and clamp cell fusion.

Dikaryons arise when both *A* and *B* are different, but heterokaryons can also be formed in matings in which one of the mating-type factors is homozygous. When *A* factors are the same (called common-*A* or *A*-off, *B*-on), nuclear migration takes place but no clamp connections form. Mating strains carrying the same *B* factor forms a heterokaryon (called common-*B* or *A*-on, *B*-off) only where the mated monokaryons meet because nuclear migration is blocked. In this case, apical cells of heterokaryotic hyphae start to make clamp connections and the nuclei divide, but the clamp (hook) cell fails to fuse with the adjacent cell, and the nucleus in the clamp remains trapped.

Detailed molecular analysis of the *A* mating-type factors show that they contain many more genes than classical genetic analysis could reveal (Fig. 2.13). In fact, each *A* locus of *C. cinereus* contains a variable number of genes, which are arranged in pairs like the *bE–bW* pair in the *U. maydis b* locus. The *C. cinereus* gene pairs were originally designated *a*, *b*, and *d*;

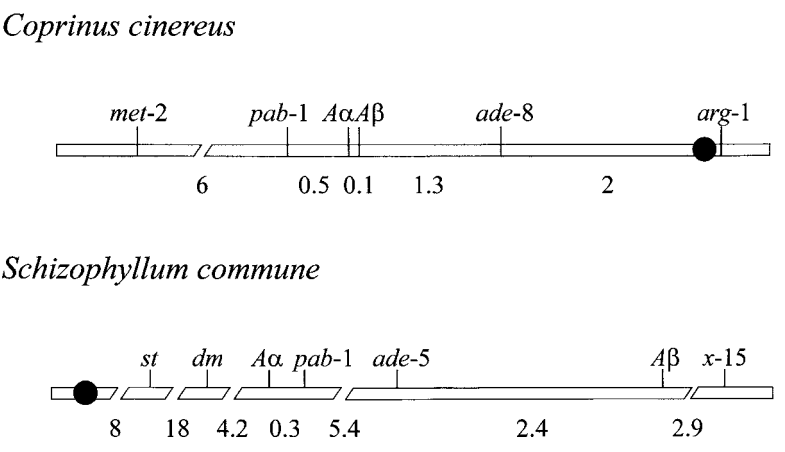


Fig. 2.12. Linkage maps of the chromosomes that carry the *A* mating-type factor in *Coprinus cinereus* and *Schizophyllum commune*. Conventional genetic analysis reveals the subloci called *A*α and *A*β. In *S. commune* these subloci are relatively far apart, but they are much closer in *C. cinereus*. The closed circles indicate the positions of the centromeres. (Figs. 2.8–2.12 adapted and redrawn from Moore (1998), *Fungal Morphogenesis*, Cambridge University Press.)

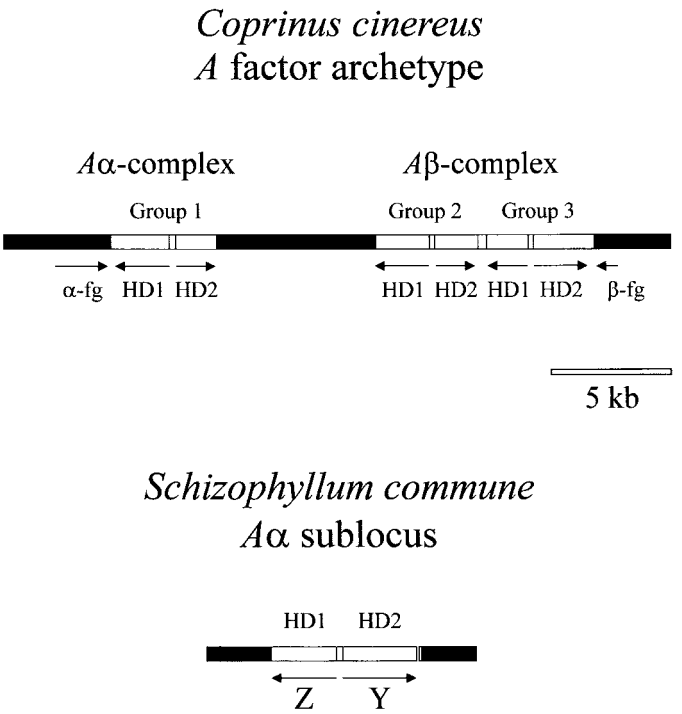


Fig. 2.13. Diagrams of parts of the *A* mating-type factors in *Coprinus cinereus* and *Schizophyllum commune*. Arrows show the direction of transcription. The (predicted) archetypal *A* factor from *Coprinus cinereus* has three pairs of functionally redundant genes (group 1, group 2, and group 3) that encode the homeodomain proteins (HD1 and HD2). The α-*fg* and β-*fg* sequences are homologous in all *A* mating-type specificities in *C. cinereus*. Interaction between HD1 and HD2 proteins is the basis of the compatible reaction (see Fig. 2.14). *A* factors examined in different strains of *C. cinereus* isolated from nature contain different combinations, and different numbers, of these genes. In *Schizophyllum commune* the mating-type genes are called *Z* and *Y* and encode HD1 and HD2, respectively. Again, different idiomorphs are found in different natural mating types; indeed, the *Z* gene is absent in the *A*1 mating type.

unfortunately, the first locus sequenced had a nonfunctional gene (which was labeled *c*) between the *b* and *d* pairs; so the *c* genes do not exist. The gene pairs have been redesignated and they are now called groups 1, 2, and 3. Each group within the *A* locus encodes two dissimilar homeodomain proteins (*HD1* and *HD2*) which are homologous to the *S. cerevisiae MAT* α*2* and *MAT a1* mating proteins, respectively.

Several features combine to ensure that there is no intragenic recombination within the *A* locus, which is likely to disturb its regular "two-by-two" structure. The groups are organized into cassettes so that they act as a single

unit and the DNA sequences are sufficiently different between groups 1, 2, and 3 to avoid homologous recombination. In addition, the paired genes are transcribed in opposite directions (Fig. 2.13). The *A* locus is bounded by DNA sequences that ***are*** homologous in all *A* mating type specificities, called α-*fg* and β-*fg*; also, the group 1 gene pair is separated from groups 2 and 3 by a 7 kbp DNA sequence that is homologous in all *A* loci (known as the "homologous hole"). The group 1 gene pair corresponds to the *A*α sublocus defined by conventional linkage analysis, whereas the group 2 and group 3 gene pairs comprise the *A*β sublocus; 7 kbp is approximately equivalent to the 0.1% recombination observed between these subloci (Fig. 2.12). These short homologous sequences limit recombination to the regions between the homeodomain loci.

The three pairs of homeodomain genes are thought to have arisen by duplication, but are now functionally independent. They are also redundant in the sense that only one compatible *HD1–HD2* gene combination is required to promote sexual development. Compatible pairs, however, must come from the same subset of genes (i.e., group 1 genes work only with group 1 genes, group 2 genes only work with group 2 genes, etc.), and providing *HD1* and *HD2* proteins come from different mating-type loci. The many *A* mating-type specificities of *C. cinereus* are derived from different combinations of alleles of the group 1, 2, and 3 genes. Only five or six alleles of each gene pair are needed to generate the estimated 160 specificities of mating-type factor *A* found in nature. There are so many functionally redundant genes in the *A* locus, however, that some A mating-type specificities in nature have lost one or more of them. Nine loci have been examined so far; only one (A_{44}) has all six genes, group 1-1 and group 3-2 are missing in mating-type A_6, and group 3-1 is missing in mating type A_5.

Basidiomycete sexual development is triggered by a dimerization between *HD*1 and *HD*2 proteins from the different *A* mating-type factors of compatible individuals (Fig. 2.14). The *N*-terminal regions of these proteins are essential for choosing a compatible partner, but not for regulating gene transcription. Sequences of different idiomorphs of the mating-type genes are dissimilar and have been interpreted as being equivalent to the highly variable region in major histocompatibility loci in mammals that forms a self–nonself recognition system. The *N*-terminal region of the mating-type protein product is the part that is essential for such self–nonself recognition. It ensures that monomers from the same mating-type idiomorph are not compatible, and that only the *hetero*dimers made between the products of the two compatible mating-type factors present in the cell are able to form the DNA-binding transcription regulators. It has also been shown that compatible protein–protein interactions (heterodimerization) are more important to compatibility at the *A* locus than is the occurrence of two homeodomains. The *HD2* homeodomain is crucial to DNA binding, but the homeodomains of *HD1* proteins are relatively dispensable.

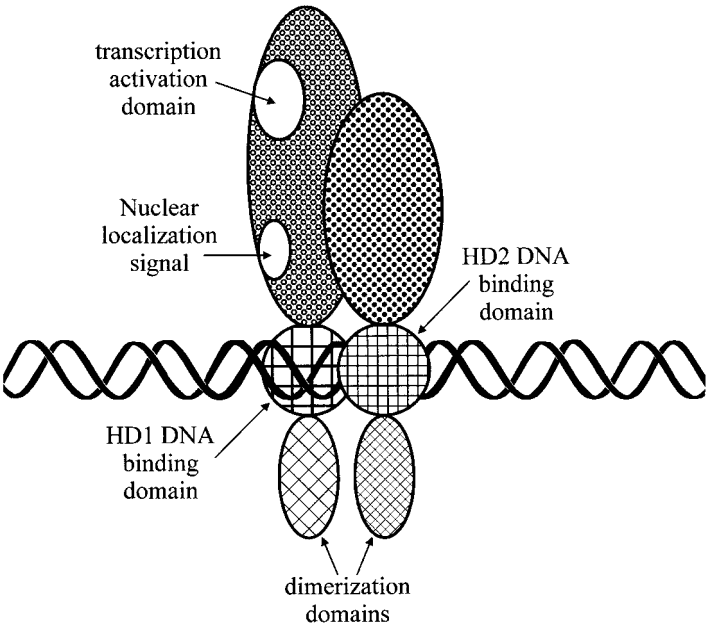


Fig. 2.14. Schematic diagram showing a model of homeodomain protein interactions involved in *A* mating-type factor activity in *Coprinus*. (Figs. 2.13 & 2.14 based on illustrations in Brown & Casselton (2001), *Trends in Genetics* 17, 393–400.)

The *A* locus of *S. commune* also controls nuclear pairing, clamp connection formation, conjugate nuclear division, and clamp septation. The *A*α locus contains the *Y* gene (which has alleles *Y*1, *Y*3, and *Y*4) and the *Z* gene (with alleles *Z*3 and *Z*4), which encode the homeodomain proteins *HD*2 and *HD*1, respectively. The *S. commune A*α locus corresponds to a single gene pair from the *C. cinereus* complex. The *A*β locus also encodes a polypeptide with a homeodomain. *Y* and *Z* are the only determinants of *A*α activity and *A*α and *A*β function independently of each other. Interactions of the *Y* and *Z* proteins have been demonstrated experimentally for nonself combinations (e.g., *Y*4 with *Z*5) proteins, but no interaction occurs between Y and Z proteins encoded by the same *A* factor (for example Y4 and Z4).

The search is under way for the genes that are regulatory targets of homeodomain proteins. Several mutants defective in mating-type-regulated events have been isolated in *S. commune* and *Ustilago*, but a consistent story has not yet emerged.

Most of the initial work on mating-type factors in *S. commune* and *C. cinereus* was aimed at determining the structure of the *A* loci. Cloning the *B* sequences revealed that the multiallelic *B* mating-type factor codes for several pheromone and receptor genes. In the smut fungi, pheromone signaling is important in cell fusion, in establishing the dikaryon, and in main-

taining filamentous growth; however, hyphal anastomosis (= hyphal fusion) of monokaryotic vegetative cells in *S. commune* and *C. cinereus* does not depend on pheromone-based recognition. As we discussed in Section 2.1, in the saprotrophic basidiomycetes hyphal anastomoses occur readily as part of the maturation process of the mycelium, and anastomosis is independent of the mating-type factors. On the other hand, it is now quite clear that pheromone signaling controls the *B*-regulated events of reciprocal nuclear migration and clamp cell fusion. Sequencing the *S. commune* *Bα1* region revealed a pheromone receptor gene (called *bar1*, standing for B-α-receptor-1) and three pheromone genes, *bap1*, *bap2*, and *bap3* (*bap* stands for B-α-pheromone). The Bβ1 locus also contains a receptor gene (*bbr1*) and genes for pheromones, *bbp1*(*1*) and *bbp1*(*2*). The *B* factor of *C. cinereus* contains 3 groups (called *1*, *2*, *3*) of genes which each code for a pheromone receptor and two pheromones. The *B* pheromone genes are all predicted to encode for lipopeptides similar to the *S. cerevisiae a* factor (see Fig. 2.5), whereas the *B* pheromone receptors are homologous to the *S. cerevisiae a* factor receptor, which is a typical G protein coupled receptor. It is worthy of note that only the *a* factor-type pheromones have been found in basidiomycetes, whereas both *a*- and α-type *receptors* are present.

The model of pheromone function that has been developed for these filamentous basidiomycetes is that after anastomosis the pheromones produced by the invading nucleus diffuse ahead and act as advance signals of nuclear migration. This activates receptors encoded by resident nuclei in nearby cells. The interactions prepare the cells for nuclear migration by initiating septal dissolution to allow nuclei to pass. *A*-factor functions then establish the dikaryotic state and clamp connection formation. Pheromone signaling is then further involved in clamp-cell fusion. There are nine mating-type specificities at the *Bα* factor of *S. commune*, each encoding a receptor and one or more pheromone(s). Hence, each receptor must distinguish at least eight nonself pheromones. Individual pheromones may also activate more than one receptor. Work with mutants in *B*-regulated functions is beginning to identify the genes subject to pheromone signaling. Some of these mutations map to the *B* loci and affect mating specificity and could be modifiers of pheromone or pheromone receptor gene specificity. Others include nine genes that influence nuclear migration. Many of these genes are linked to *B*, which suggests that related functions are clustered.

2.10 Overview: Biology of Incompatibility Factors

Mating-type factors regulate pheromone and pheromone receptors involved in mating. This might be interpreted as ranging from recognition between sexually competent cells in yeast to governing growth of clamp connections in basidiomycetes. It has also been suggested that *B*-mating-type pheromone signaling is involved in internuclear recognition, and, in

particular, regulation of the distance between the two nuclei in the dikaryon. Furthermore, heterodimerization of homeodomains from different idiomorphs is employed to transcriptionally regulate further aspects of sexual development.

In the highest expression of this activity, compatible mating-type factors permit the development of complex mushroom fruit bodies, which contain several different interacting tissues and, in some tropical genera, may have caps approaching 1 m in diameter. Not all fungi, however, possess mating-type factors. In *Podospora*, the progress of meiosis and sporulation does not require heterozygous mating-type factors. Indeed, apparently normal fruiting bodies can be formed by haploid cultures, and fruit body formation can usually be separated from other parts of the sexual pathway by mutation, even in species that have a well-developed mating-type system. The significance of mating-type factors in regulating events beyond the initial mating reaction, therefore, is difficult to judge.

Vegetative compatibility genes define the individuals of fungal populations, and mating-type factors are usually interpreted as favoring the outbreeding of a fungal population. Only 25% of siblings are able to mate with a bifactorial mating-type system, so it favors outbreeding. As a result, mating-type genes contribute to management of the genetics of the population as well as to the sexual development of the individual. Sexual reproduction generates genetic variation, offers an escape from DNA parasites, and provides a means to repair DNA damage. Sexual reproduction is an important way of enhancing the overall rate of adaptation of the species. This is despite the fact that most fungi produce asexual spores that are extremely effective in dispersing the organism. These are usually produced in such very large numbers that even small quantities of substrate might be expected to produce a sufficient number for mutation alone to provide the variation on which selection might operate. If this can be the case, then we have to ask why so many fungi invest more resources in sexual reproduction.

There are, admittedly, many fungi that only reproduce asexually, but the majority does have a sexual cycle and this must have a selective advantage. Asexual stages would otherwise replace sexual ones entirely. Because mitotic nuclear divisions multiply the genetic constitution of only one individual in asexual reproduction, the crucial, and contrasting, step in sexual reproduction is the fusion of nuclei derived from different individuals. If the individuals differ in genotype, the fusion nucleus will become heterozygous and the products of the meiotic division (= the progeny if the organism is haploid, or the gametes if the organism is diploid) can have recombinant genotypes. Thus, in one sexual cycle, new combinations of characters can be created in the next generation for selection. This is the most usual "explanation" for sex; namely, that it promotes genetic variability through outcrossing and that variability is needed for the species to evolve to deal with competitors and environmental changes. There is plenty

of evidence to show that asexual lineages change little in time and that outcrossing certainly does promote variability in a population, which enables the organism to survive ecological and environmental challenges.

This, though, is a "group selectionist" interpretation. It argues that variation generated in an *individual* meiosis benefits the *group* or population to which the individual belongs. Current theory emphasizes, however, that selection acts on individuals, so any feature that is advantageous in selection must be so because it benefits either the individual itself or its immediate progeny. An alternative interpretation suggests that repair of damaged DNA is the crucial advantage of the meiotic sexual cycle. DNA damage in one chromosome, caused by mutation or faulty replication, can be repaired by comparison and recombination with the normal chromosome provided by the other parent. Genetic fitness would be increased, but only when outcrossing ensures heterozygosis. Even an incomplete sexual cycle might be of advantage in this case.

Mutations can be recessive and damaging, and different mutations will occur in different mitotically generated cell lines. Merely the formation of the diploid (or heterokaryon or dikaryon) by outcrossing will benefit the mated individual if recessive adverse mutations are masked by nonmutant alleles in the nuclei of the other parent. Outcrossing might also give rise to heterozygous advantage, where the heterozygous phenotype is better than either of its homozygous parents, which has frequently been demonstrated in plants and animals and has also been demonstrated in *Saccharomyces cerevisiae*. These alleged advantages of the sexual cycle are not mutually exclusive, nor of equal value; rather, they are themselves phenotypic characters that may or may not have selective value for the organism concerned. Different species have different life cycles and experience different evolutionary challenges; therefore, they may make use of, enhance, or dispense with various aspects of sexual reproductive processes for any one or more than one of the interpretations outlined herein. Life is a rich tapestry.

Publications and Websites Worth a Visit

Bartholomew, K.A., Marion, A.L., Novotny, C.P. & Ullrich, R.C. (1996). A case study in fungal development and genetics: *Schizophyllum commune*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 371–384. Marcel Dekker, Inc.: New York.

Bölker, M. (1998). Sex and crime: heterotrimeric G proteins in fungal mating and pathogenesis. *Fungal Genetics and Biology* **25**, 143–156.

Bloom, K. (2001). Nuclear migration: cortical anchors for cytoplasmic dynein. *Current Biology* **11**, R326–R329.

Brown, A.J. & Casselton, L.A. (2001). Mating in mushrooms: increasing the chances but prolonging the affair. *Trends in Genetics* **17**, 393–400.

Casselton, L.A. & Olesnicky, N.S. (1998). Molecular genetics of mating recognition in basidiomycete fungi. *Microbiology and Molecular Biology Reviews* **62**, 55–70.

Dalgaard, J.Z. & Klar, A.J.S. (2001). Does *S. pombe* exploit the intrinsic asymmetry of DNA synthesis to imprint daughter cells for mating-type switching? *Trends in Genetics* **17**, 153–157.

Debuchy, R. (1999). Internuclear recognition: a possible connection between euascomycetes and homobasidiomycetes. *Fungal Genetics and Biology* **27**, 218–223.

Elliott, C.G. (1994). *Reproduction in Fungi. Genetical and Physiological Aspects*. Chapman & Hall: London, UK.

Glass, N.L., Jacobson, D.J. & Shiu, P.K.T. (2000). The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annual Review of Genetics* **34**, 165–186.

Haber, J.E. (1998). A locus control region regulates yeast recombination. *Trends in Genetics* **14**, 317–321.

Hegner, J., Siebert-Bartholmei, C. & Kothe, E. (1999). Ligand recognition in multiallelic pheromone receptors from the basidiomycete *Schizophyllum commune* studied in yeast. *Fungal Genetics and Biology* **26**, 190–197.

Kothe, E. (1999). Mating types and pheromone recognition in the homobasidiomycete *Schizophyllum commune*. *Fungal Genetics and Biology* **27**, 146–152.

Kronstad, J.W. & Staben, C. (1997). Mating type in filamentous fungi. *Annual Review of Genetics* **31**, 245–276.

Shiu, P.K.T. & Glass, N.L. (2000). Cell and nuclear recognition mechanisms mediated by mating type in filamentous ascomycetes. *Current Opinion in Microbiology* **3**, 183–188.

Historical Publications Worth Knowing About

Blakeslee, A.F. (1904). Sexual reproduction in the Mucorineae. *Proceedings of the American Academy of Arts and Sciences* **40**, 205–319.

Buller, A.H.R. (1931). *Researches on Fungi*, vol. 4. Longmans: London, U.K.

Buller, A.H.R. (1933). *Researches on Fungi*, vol. 5. Longmans: London, U.K.

Buller, A.H.R. (1934). *Researches on Fungi*, vol. 6. Longmans: London, U.K.

Raper, J.R. (1939). Sexual hormones in *Achlya*, I. Indicative evidence for a hormonal coordinating mechanism. *American Journal of Botany* **26**, 639–650.

Sansome, E. (1961). Meiosis in the oogonium and antheridium of *Pythium bebaryanum. Nature* **191**, 827–828.

CHAPTER 3

Wild Types and Mutants

Revision Concepts

- Mendel's first law of inheritance is that two alleles of a gene (= two homologous chromosomes) segregate from each other into the gametes.
- Mendel's second law of inheritance is that alleles of different genes segregate independently.
- Haploidy in microorganisms allows the phenotype to represent the genotype directly.
- Advantages of using fungi as a model include the production of millions of haploid progeny and the ability of many to grow on simple mineral-based (minimal) media.
- Mutations can range from the simplest type (one base pair or point mutations) to large mutational changes, consisting of several base pairs to large segments of the chromosome, or deletions.
- Inversions result when the segment between two breaks is reinserted in reversed orientation in the same chromosome, whereas translocations result when a segment is removed and transferred into another chromosome.
- Strains containing less than one full set of homologous chromosomes are called *aneuploids*; those containing more than one set are called *polyploids* (not common in fungi).
- Karyotyping is the assessment of the variation in chromosome number and size.
- Natural spontaneous mutation frequency is low.
- Mutation *via* X-ray radiation results in deletions and inversions, which require two breaks, but also DNA base mispairing and base substitutions.
- Ultraviolet radiation (UV) does not cause two breaks, but it can cause dimerization between adjacent pyrimidine bases; UV mainly causes G–C to A–T transition mutations.
- Different chemical mutagens bring about G–C to A–T transitions, transversions, and frameshift mutations.

- DNA glycolyases excise chemically modified bases, but more substantial damage to DNA is dealt with by the excision repair system.
- Auxotrophs are stable nutritional mutants that can only grow on minimal medium if it is supplemented with the metabolite they cannot synthesize.
- The auxanographic technique (sprinkling crystals of potential nutrients onto the medium) is one of several ways of identifying the specific growth requirements of auxotrophs.
- Techniques to enhance isolation of auxotrophic mutants include filtration enrichment and replica plating.
- Resistance mutants grow on media on which parental strains fail to grow.
- *Intra*genic suppression re-establishes the correct reading frame by a second mutation in the same gene that already carries a frameshift mutation.
- *Inter*genic suppression is caused by a second mutation in a different gene, often a tRNA gene.
- Restriction fragment length polymorphisms (RFLPs) are pieces of DNA produced by digestion with a restriction enzyme, which cleaves at specific sequences to produce a restriction pattern or restriction phenotype.
- RFLPs can be used to distinguish genotypes, to construct genetic maps, as a marker to other traits, or to fingerprint a particular individual.
- Site-directed mutagenesis functions by disrupting a particular gene with unrelated, transforming DNA that has enough homology with the target gene to locate it on the chromosome.
- Site-directed mutagenesis can be used to study gene function, but prior knowledge of the DNA sequence of interest is required.
- Gene disruption using transformation and homologous recombination can be used to obtain mutants in pathogenicity, morphogenesis, and development.

3.1 Phenotypes and Genotypes: Variations and Mutations

Genetic analysis of any organism requires genetic variation. The originator of the science, Gregor Mendel, recognized this, and the two features of his work that made it so revolutionary were the fact that he counted progeny, and he used heritable characteristics that had two contrasting phenotypes.

Geneticists need to count progeny and manipulate numbers with statistical precision. We will deal with these aspects later in Sections 4.8 and 4.9. For the moment we will concentrate on the variation. Mendel's use of contrasting phenotypes in his pea plants, like tall versus short stems or smooth versus wrinkled seeds, enabled him to recognize the genetic basis of the segregations he was dealing with and postulate that each phenotype is specified by a different allele of the gene determining that phenotype. This interpretation of the segregations in terms of dominant and recessive alleles existing in pairs in a diploid was of course only a theory as far as Mendel

was concerned (and the word *gene* was not coined until long after his death), but we now know it to be fact.

Mendel, and those early researchers who immediately followed him, used phenotypes that were distinguishable by eye (i.e., morphological characters like stature, shape, color, and the like). This is obviously a limitation, particularly so with fungi because fungal mycelia have so few distinctive morphological features with which to start. The approach, however, gave Mendel the opportunity to establish his two Laws of Inheritance: the First Law, that the two alleles of a gene segregate from each other into the gametes, and the Second Law, that alleles of different genes segregate independently. These are now understood to represent the behavior of chromosomes in meiosis. The two alleles of a gene are located on the two homologous chromosomes in a diploid nucleus, and Mendelian segregation results from the way in which meiosis produces the four haploid daughter nuclei (which enter gametes in diploid organisms, but spores in fungi that are haploid in their growth phase).

Phenomena not observed in Mendel's classic segregations, such as partial dominance, codominance, gene interactions, and linkage (when the segregating genes are located on the same chromosome), were all discovered in the early years of the twentieth century using morphological characters. The limitations of using morphological phenotypes were recognized at about the time that the concept of gene-directed metabolism was emerging. As the truth of this concept became established, attention turned to genetic characteristics that were biochemical or metabolic in nature. Indeed, the biochemical phenotypes of *nutritional mutants* of bacteria and fungi actually established much of what we know about metabolic pathways.

Microbes that can be readily grown on simple media in the laboratory are more attractive as experimental organisms than others that may be seasonal (e.g., crop plants), or those that may require extensive housing (e.g., animals or even insects), or those that may produce only a limited number of offspring from each cross (e.g., most higher animals). Haploidy is also an advantage, of course, because the phenotype directly represents the genotype in a haploid, with none of the complications caused by dominance that occur in diploids. Production of large numbers, potentially many millions, of haploid spores that are easily grown makes microbes excellent candidates for studies of even rare genetic events. The important difference from other organisms, though, is the fact that many microbes, the saprotrophs in general, grow well on a simple medium, which, providing it contains an organic source of carbon and a few vitamins, can otherwise be composed of totally inorganic constituents. Even the nitrogen source can be an inorganic saltlike sodium nitrate or ammonium chloride. Such an organism is called a *prototroph* because it feeds on minimal medium using the simplest of nutrients. At some stage in its history the laboratory prototroph will have been derived from an isolate collected from its natural environment: this strain is called a *wild type*. All of this applies equally to bacteria and fungi,

of course, but fungi are eukaryotes and this combination of features raises the significance of fungal biochemical genetics to a higher level altogether. Fungi provide the experimental means by which we can gain a better insight into the workings of all other eukaryotes.

An organism growing on a largely mineral medium evidently must be able to synthesize all of the organic components of the cell for itself. If there are no amino acids supplied in the medium, the organism must be synthesizing the amino acids for its proteins, as well as the purine and pyrimidine bases for its nucleic acids and energy metabolism, and all of the other building blocks of the cell. Now, the argument runs like this. If the wild type is able to synthesize everything it needs because it possesses a set of genes that provide all the enzymes involved in the numerous pathways of normal cell metabolism, then the experimenter should be able to isolate nutritional mutants, strains that have lost one or another synthetic capability because of a defect in a gene controlling that capability. These strains are called *auxotrophs* because they have a nutritional requirement; they can only grow if the minimal medium is supplemented with the metabolite that they cannot synthesize.

Variant strains like this arise through abrupt and stable changes in the DNA called *mutations*. Fungal populations in nature (wild populations) usually contain naturally occurring allelic differences because mutations occur spontaneously in all populations. Such natural variants provide good material for studies of molecular polymorphisms, where the method of analysis may be able to detect differences as small as changes in single base pairs; however, these often have no effect, or too small an effect, on expression of the phenotype to be useful for conventional genetic experiments. They lack the sharp contrast in allelic phenotypes that make genetic investigation possible. Nutritional mutants, on the other hand, cannot grow on minimal medium, but they can grow on minimal medium supplemented with the required nutrient. A growth–no growth contrast like this is ideal for research. Even though most mutants of this sort could not survive in nature, they make very good experimental material. It is useful to distinguish between forward and reverse mutations. Mutations from the wild type to a mutant condition are *forward mutations*, whereas mutations from a mutant phenotype back to, or close to, the wild type are *reverse mutations*.

3.2 Molecular Nature of Mutation

Substitution of one base pair with another is the simplest type of mutation; it is often called a *point mutation*. A change of a purine for the other purine and a pyrimidine for the other pyrimidine is called a *transition*, whereas the term *transversion* is applied when a purine replaces a pyrimidine or vice versa. Because the genetic code is degenerate, many transitions and some transversions have no effect on the polypeptide gene product. This is

because the mutation either produces another codon for the same amino acid or because when an amino acid replacement occurs, the new amino acid has chemical characteristics sufficiently similar to the original one for there to be no change in protein function. Such cryptic mutations are clearly not detectable as functional mutants, but they are detectable when molecular techniques of mutant characterization are used.

When one or a few base pairs are inserted or deleted, the reading-frame for translation of the messenger RNA is altered. Insertions and deletions of any number of base pairs other than three or a multiple of three will cause such frame-shift mutations. The change of reading frame leads to incorrect translation of the message after the mutant site, and a completely wrong amino acid sequence will be produced that covers either the entire sequence after the mutation or until a termination codon is produced by the frame-shifting. On average, you would expect translation of the wrong-reading frame to persist for about 21 codons before termination occurs. This is calculated from the fact that because 3 of the 64 codons are chain-terminators, they will occur with a frequency of 3 in 64 (= 1 in 21) in a random sequence in which the four bases occur in equal numbers.

Deletions are larger mutational changes that vary in size from several base pairs up to large segments of the chromosome. If they occur within a gene a much-altered product (if any) will be formed, but the larger deletions can cover several genes and consequently eliminate several functions. Such major changes are most likely to be lethal in haploids. Addition of DNA segments, called *duplications*, also occurs. Adding genes may not alter the phenotype, but when the duplication occurs it may be inserted into the chromosome within some other gene, which will usually eliminate that gene function.

Drastic changes in chromosome structure, and even chromosome number, can also occur. Chromosome breakage can occur spontaneously or in response to mutagens, and several types of chromosomal rearrangements can result. Inversions result when two breaks occur in the same chromosome and the segment between them is reinserted in reversed orientation. An inversion that includes the centromere is called *pericentric*; inversions limited to one chromosome arm are called *paracentric*. A chromosome segment may be removed from its normal chromosome and transferred into another chromosome; these are *translocations*. They are often reciprocal, happening when arms of different chromosomes exchange segments, but insertions sometimes occur in which a chromosome segment from one chromosome is inserted at a break site of another (a *transposition* is the insertion of a segment into another part of the same chromosome). Structural rearrangements of these sorts cause altered linkage relationships, and the most drastic ones cause inviability in the products of meiosis of heterozygotes. Strains with duplicated chromosome segments are often genetically unstable; this is expressed by the formation of more vigorously growing sectors from colonies that comprise hyphae containing nuclei from which

most or all of the duplicated material has been deleted, usually from the structurally abnormal chromosome.

Strains containing more than two whole normal sets of chromosomes are called *polyploids*; *autopolyploids* when the chromosome sets come from the same species and are homologous with each other or allopolyploids if chromosome sets from different species have hybridized. Haploids have one set of chromosomes [symbolized n, diploids two sets ($2n$), and then there are various polyploids such as triploids ($3n$) with three sets, and tetraploids ($4n$) with four]. Polyploidy is very common in plants, but much less so in fungi. Triploids can be made in *Saccharomyces cerevisiae* by crossing a diploid homozygous for one mating type with a haploid of the opposite mating type. Meiotic products of triploids always have reduced viability because the abnormal 2:1 disjunction of the three homologous chromosomes (so-called trivalents) at the first division of meiosis produces many nuclei with aberrant chromosome numbers that do not survive. Autotetraploids have a more regular chromosome disjunction and can become established, although the frequencies of genotypes expected among the progeny can be quite complex as there are, for example, three kinds of heterozygote for each gene.

Aneuploids have less than a whole haploid set of extra chromosomes. The simplest have just one chromosome less or one chromosome more than the diploid or haploid [e.g., trisomics ($2n + 1$) and monosomics ($2n - 1$), or the disomic ($n + 1$)]. Aneuploids arise through disjunction errors in mitosis or meiosis. Aneuploidy has been studied particularly in *Aspergillus nidulans* where aneuploid nuclei are formed as synthetic diploids undergo haploidization as part of the parasexual cycle (see Section 5.10). *Saccharomyces cerevisiae*, however, is unusually tolerant of anything up to five extra chromosomes. Artificial yeast chromosomes (called YACs) have been made up from yeast centromere DNA, yeast telomeres, and yeast replication origins to serve as large capacity cloning vectors for genetic engineering (see Section 8.5).

Assessment of variation in chromosome number and size, called *karyotyping*, can be done using electrophoresis systems that separate DNA molecules of chromosomal size, that is up to two megabases (2Mb). The technique reveals chromosome length polymorphisms (CLPs) in many fungi (see Section 7.12, and Fig. 7.2), which can arise from chromosome rearrangements like translocations. Pathogenic fungi may have dispensable or supernumerary B chromosomes. These tend to be small in size, generally do not contain functional genes, and are unstable during meiosis.

3.3 Natural Variation and Spontaneous Mutation

A difficulty with natural variation is that although it is unavoidable, spontaneous mutation occurs at very low frequency. Spontaneous mutation frequency can be determined from the proportion of mutant spores produced

by a series of replicate cultures that were started with small inocula. If the sample size is large enough the cultures will contain the population extremes: those containing many mutant spores because the mutations occurred early in growth and cultures containing unusually few mutant spores because the mutations occurred late in growth. The majority of the cultures will contain much the same, fairly low, number of mutants, and these will give an approximation of the spontaneous mutation frequency. This is always likely to be overestimated because of differences in selective advantage of mutations and the timing of their occurrence during growth of the test cultures.

Experiments have shown that spontaneous mutation occurs in dry spores during storage; it's a mistake to think of it occurring only in dividing nuclei. The spontaneous rate of forward mutation is usually in the region of 10^{-6} per gene. In nondividing cells, spontaneous mutations are presumably caused by spontaneous chemical changes to DNA that might be enhanced by elevated temperatures, background radiation, and cosmic rays. In active cells, errors during DNA replication can generate spontaneous mutations. These errors include abnormal base pairing by unusual tautomeric forms of DNA bases causing substitutions of one base pair for another, and mispairing of DNA strands (particularly where the base sequence is repeated) leading to the addition or deletion of one or a few base pairs.

3.4 Generating Mutants: Mutagenesis

A natural mutation frequency of 10^{-6} per gene is not high enough to provide all the mutants a genetical study will require; thus, treatments must be used to increase the mutation rate to a level at which mutants can be efficiently isolated and characterized. These mutagenic treatments include irradiation with ultraviolet light or with X-rays, and a wide variety of chemical treatments. We will discuss some of the treatments later. The general approach, however, is to treat spores (which can be in suspension in sterile water) with the mutagen. The extent of the treatment depends on the strength of the mutagen and the sensitivity of the spores. Treatments that increase the mutation rate to a useful extent also kill many of the treated spores. As a result a compromise must be established between having too few spores surviving (at a high mutagen dose) and having many survivors but with very few of them being mutants (at a low dose).

Pilot experiments must therefore be carried out to determine the amount of kill for a range of mutagen exposures by plating untreated and treated samples, after dilution, on to a complete medium (containing every kind of nutrient that auxotrophs might be expected to require). "Survival curves" can then be drawn relating viability to the extent of mutagen exposure. In general, as the dose of mutagen is increased, the yield of mutants increases,

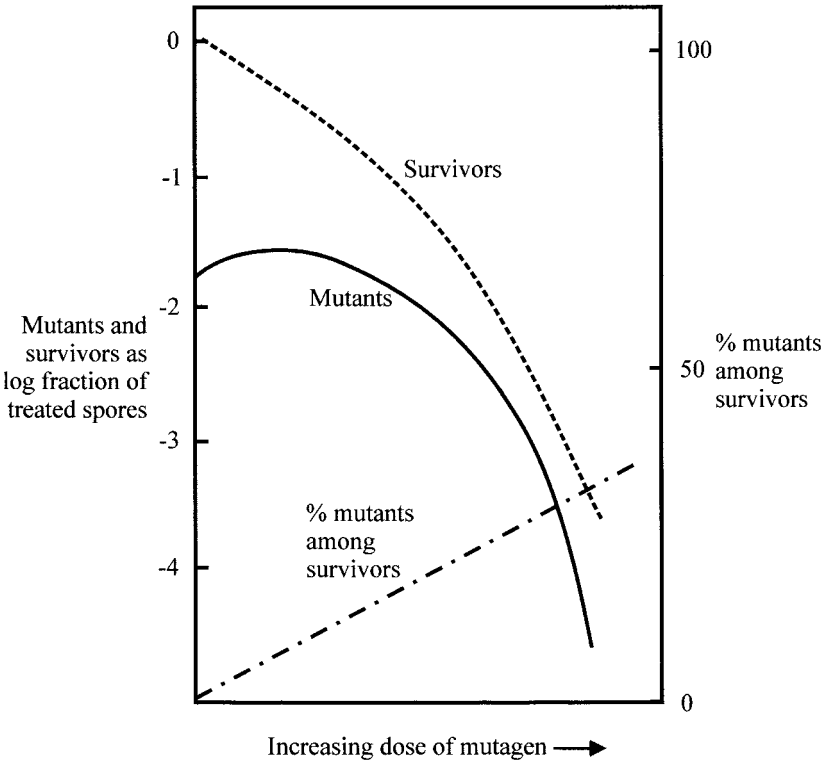


Fig. 3.1. Representative mutation kinetics curves. As the dose of mutagen is increased, the yield of mutants increases but the number of survivors decreases. The exact shape of these curves depends on the organism, mutagen, and treatment conditions. (Adapted and redrawn from Fig. 61(a) in Fincham, Day & Radford (1979), *Fungal Genetics*, Fourth edition, Blackwell Scientific Publications.)

but the number of survivors decreases. A very generalized set of mutation kinetics curves is shown in Fig. 3.1.

The yield of mutants often shows a greater than linear increase with mutagen dose over the lower dosages. In some cases the proportion of mutants among survivors can decrease if repair mechanisms are induced. Detailed survival curves can be used to study mutation kinetics, but treatments that give 95–99% kills can be used routinely for mutant hunts. Surviving spores can be plated onto complete medium, with the dilution factor being reduced to take account of the expected kill. Colonies that grow from these platings can then be subcultured into individual slopes prior to individual testing of their phenotypes to search for nutritionally defective mutants. It is a great deal easier to use a selective technique to isolate mutants when feasible; for example, mutants resistant to some growth-inhibiting supplement. Very large numbers of spores can be plated in such

selective procedures because the bulk of them are not expected to grow. For example, oidia of *Coprinus cinereus* have been plated at about 10^5 on each 9 cm diameter Petri dish during experiments to isolate mutants resistant to inhibition by sugar analogues. With such large numbers of spores it is best to use "pour plates" in which the spores are mixed with the molten medium (cooled to about 50°C) instead of being spread over the agar surface before the plates are poured.

All bases in the sequence of a gene are candidates for mutation, but they don't all mutate with equal frequency. There are regions, known as *mutational hot spots*, that are highly sensitive to mutagenesis. The location of such a hot spot will depend on the mutagen in use at the time. Different mutagens show different site preferences, presumably because neighboring bases influence the chemical environment of the base pair that actually mutates.

3.5 Generating Mutants: Irradiation

Radiation can damage DNA in a variety of ways. X-rays can break the phosphodiester backbone, resulting in deletions and inversions. Energy from the ionization caused by X-rays can also produce local chemical conditions able to deaminate and dehydroxylate DNA bases leading to mispairing and base substitutions. In general for fungi conditions can be found in which the mutation frequency is proportional to X-ray dose, which suggests that each mutation is induced by a single ionization track. The mutations, therefore, are likely to result from base substitutions because deletions require two breaks; thus, the frequency of deletions is, in the ideal case, proportional to dose squared (Fig. 3.2).

Ultraviolet light (UV) covers a range of wavelengths just shorter than the visible spectrum. It has lower energy than do X-rays and does not cause ionization; therefore, it doesn't break the phosphodiester backbone. The absorption spectra of purines and pyrimidines extend into the UV range, however, so UV is strongly mutagenic to these DNA bases. The main effect is the covalent linking of adjacent pyrimidine residues into pyrimidine dimers, which can occur wherever two pyrimidines occur next to one another in the DNA sequence. Thymine dimerizes more readily than cytosine, and about half the dimers formed are thymine–thymine dimers. Covalent links between adjacent bases are errors in DNA structure, so pyrimidine dimers must be removed from DNA before it can replicate. There is a widely distributed enzyme called DNA photolyase that uses the energy of visible light to uncouple the dimers and restore the original DNA structure. This is called *photoreactivation*. If the dimer persists, excision-repair can take place in which an endonuclease excises a segment of the DNA strand adjacent to the dimer, creating a single-stranded gap. The gap

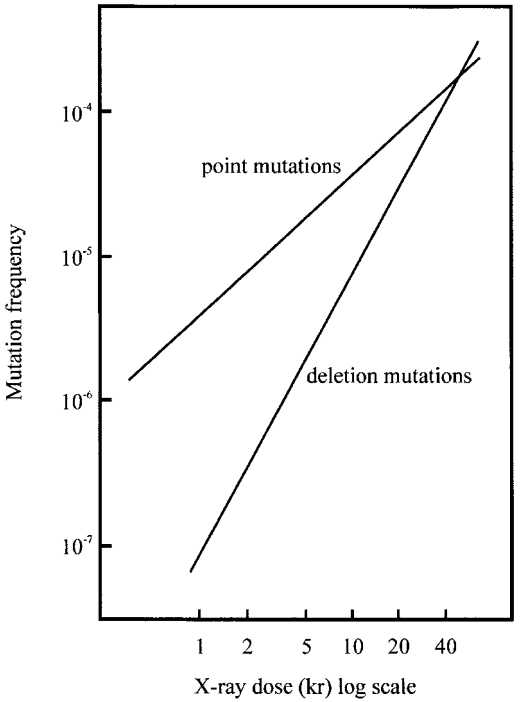


Fig. 3.2. Logarithmic plots of point mutation and deletion mutation frequencies related to X-ray dose in conidia of *Neurospora crassa*. Mutation frequency is proportional to dose, which suggests that each mutation is induced by a single ionization track. Deletions require two breaks, so the frequency of deletions is proportional to the square of the dose. (Adapted and redrawn from Fig. 61(b) in Fincham, Day & Radford (1979), *Fungal Genetics*, Fourth edition, Blackwell Scientific Publications.)

is repaired by new synthesis in which the DNA polymerase uses the intact complementary DNA strand as a template (see Section 6.4).

UV mutants result from errors in these repair pathways, which are normally involved in DNA replication and recombination. They seem to be error-prone repair pathways. Mutations have been isolated from several fungi that have enhanced sensitivity to mutagens as a result of defects in the repair pathway. Mutations in components of these pathways also have been isolated that exhibit mutagen resistance because they have repair pathways that are more efficient than the wild type. The logic of this seems to be that under normal conditions (i.e., without the experimenter's imposed mutagen treatment), error-prone repair pathways provide a means of generating the spontaneous variants that are tested for evolutionary

fitness in the natural population as they go about their normal business in replication or recombination. The efficiency of the DNA repair pathways is evidently open itself to evolutionary adaptation.

In addition to forming pyrimidine dimers, UV light can also form cytosine hydrate, which may pair with adenine at the next replication so that the wild-type G–C base pair becomes a mutant A–T base pair. Further, dimerized cytosine tends to deaminate to uracil and can lead also to the G–C to A–T transition if photoreactivated. UV consequently causes mainly G–C to A–T transitions, although up to 50% of the mutants will be insertions and deletions caused by error-prone repair of pyrimidine dimers. The relation between mutation frequency and UV exposure is often complex because of the involvement of repair pathways. UV induces the repair system responsible for most UV-induced mutation. This can even result in a decrease of mutation frequency with exposure because of enhanced repair activity at higher UV doses.

3.6 Generating Mutants: Chemical Mutagens

Nitrous acid (HNO_2) is probably the simplest chemical mutagen. It is easily produced by dissolving sodium nitrite in a solution buffered at pH 4–6. Nitrous acid hydroxylates primary amino groups, and by so doing converts adenine to hypoxanthine (which pairs like guanine), cytosine to uracil (which pairs like thymine), and guanine to xanthine (no change in base pairing). As a result nitrous acid induces base-pair transitions in both directions (A–T to G–C and G–C to A–T). There is often a preference for G–C to A–T transitions, however, because the surrounding base sequence affects the reactivity of a particular base.

Hydroxylamine (NH_2OH) reacts only with cytosine (the chemical inserts itself across the carbon to nitrogen double bond in the pyrimidine ring) and brings about G–C to A–T base pair transitions. Methoxylamine (*O*-methylhydroxylamine) is more effective than hydroxylamine, but it has the same highly specific mutagenic action.

Alkylating agents replace a hydrogen atom in a compound with an alkyl group (which is a saturated open-chain hydrocarbon), and many are effective as mutagens. The most commonly used alkylating agents have methyl or ethyl as the alkyl group; they include methylmethanesulphonate (MMS, $CH_3SO_2CH_3$), ethylmethanesulphonate (EMS,) and *N*-methyl-*N*′-nitro-*N*-nitrosoguanidine (MNNG, also called nitrosoguanidine or NTG, $(CH_3NNO)(CH_3HNNO_2)C{=}NH$), all of which have been used extensively as potent mutagens with fungi. The mutagenically effective targets for alkylations are N-3 of adenine and guanine, O-6 of guanine, O-4 of thymine, and N-4 of cytosine. Alkylation of a DNA base changes its pairing relationships and may make it a substrate for the error-prone repair system.

Alkylating agents are especially effective in inducing G–C to A–T transitions, but induce all kinds of mutations, and some, such as the nitrogen mustards (chloroalkyl amines like *bis*(2-chloroethyl)ethylamine, $CH_3CH_2N(CH_2CH_2Cl)_2$), are effective in inducing transversions.

DNA repair systems can also recognize and act on DNA damage caused by chemical mutagens. They search for abnormalities in DNA structure and make appropriate chemical changes to restore the structure to normal. Many organisms have enzymes able to repair alkylation damage by transferring the alkyl group from the nucleotide to the enzyme protein. Chemically modified bases can be excised by DNA glycolyases, which swing the damaged base out of the DNA helix and break the glycosidic bond between the base and the sugar.

Each glycolyase has a limited specificity. For example, most organisms have glycolyases specific for deaminated bases like uracil (deaminated cytosine) and hypoxanthine (deaminated adenine), oxidation products such as 5-hydroxycytosine and thymine glycol, and methylated bases such as 7-methylguanine and 2-methylcytosine. Removal of the damaged base creates a baseless site that is converted into a single nucleotide gap and subsequently filled by DNA polymerase. After gap filling, the final phosphodiester bond is completed by DNA ligase. More extensive damage, causing distortion of the double helix, is repaired by the excision-repair system (see Section 6.4).

Acridine, or dibenzopyridine, is the parent compound of a group of drugs and dyes that have been used as antiseptics, but were recognized as mutagens of bacteriophage, in 1961, in which they cause frameshift mutations. The acridine double-ring does this by intercalating in the stack of base pairs in double-stranded DNA. Disruption of the strands results in the insertion or deletion of a base (frameshift mutations) as an accompaniment of recombination or as an error in replication.

Chemical analogues of purine and pyrimidine bases can also cause mutations by being incorporated into DNA during replication in place of normal bases. Two widely used base analogues are 5-bromouracil (5BU), which can be incorporated in place of thymine, and 2-aminopurine (2AP), which can substitute for adenine. Their mutagenic effect depends on their ambiguous pairing properties during DNA replication. For example, 5BU incorporated as thymine can pair like cytosine: an A–T to G–C transition is the result. Uptake of base analogues by fungi is inefficient (i.e., the uptake systems have a high affinity for the normal molecules) and their incorporation into DNA is usually limited as a result. Another interesting base analogue is caffeine (which is 1,3,7-trimethyl-2, 6-dioxo-purine). In many organisms, caffeine causes chromosome breakage and it is mutagenic to the fungus *Ophiostoma*; however, you would have to drink enough coffee to refloat the Titanic to suffer any ill effects yourself!

That facetious comment reminds us that before we leave the topic of mutagens we must emphasize that all DNA is open to attack, and that

means the experimenter's DNA as well as the fungal DNA. Extreme care must be exercised at all times when handling mutagens because they are extremely hazardous carcinogens. We have referenced a useful website for information about carcinogens at the end of this chapter.

3.7 Types of Functional Mutant

Phenotypes of forward mutations include morphological mutants in which the mutation results in changes in color, hyphal growth pattern (changing the morphology of the colony), or fruit body shape, biochemical mutants, which lead to nutritional deficiencies by removing or reducing the ability to carry out specific metabolic reactions, and resistance mutations that confer tolerance to analogues of metabolites, antibiotics, or heavy metals that are normally inhibitory to the wild type.

Some morphological mutants are particularly useful for experimental work. Spore color mutations, for example, are especially valuable as genetic markers because of their ease of scoring; the results are immediately visible in the color of the progeny spores (see Chapters 4–6 for examples). Another morphological mutant phenotype commonly isolated is one showing vigorous but abnormally compact hyphal growth. These are called *colonial mutants*. The wild type of *Neurospora crassa* has fast-growing and invasive hyphae. Much use has been made of colonial mutants in studies of *Neurospora* genetics for the simple reason that their more restricted growth pattern enables a larger number of colonies to be tested on standard-sized (9 cm diameter) Petri dishes. Colonials of other fungi have also been used in particular experimental approaches where the comparison between the widely spreading wild-type growth habit and a compact, densely branched colonial habit provides a useful visual result for an experiment. Morphological mutations are of limited use as markers in genetic analysis, however, because it is often difficult to score multiple morphological mutations in recombinant progeny.

Aside from their usefulness as experimental tools, morphological mutants themselves are of course intrinsically interesting for the light their study can shed on hyphal, tissue, and spore differentiation, as well as on fruit body morphogenesis (see Chapter 10). Morphological mutants of fungi are not greatly different in conceptual terms from morphological mutants in peas, fruit flies, or mice. The greatest advance in eukaryote genetics arguably resulted from the use of biochemical mutants, which were first studied in *Neurospora* in 1945 by G.W. Beadle and E.L. Tatum. Auxotrophs unable to grow on minimal medium have been the most widely used, although, of course, morphological and resistance mutants must also have a biochemical basis.

Mutants we have described so far grow on complete medium, and it is eventually possible in most cases to identify their exact nutritional require-

ment so that they can be cultivated on a minimal medium supplemented with the compound for which they are auxotrophic; however, a large proportion of mutations induced by mutagenic treatments are lethal. In diploid organisms, techniques have been developed to preserve recessive lethal mutations in heterozygous form. Such mutations are always lethal in haploid fungal spores, but they can survive in a heterokaryon. Experiments with *Neurospora crassa* heterokaryons indicate that auxotrophic mutants that respond to the addition of a single substance to minimal medium, and thus result from simple metabolic lesions are much less frequent than the lethal mutants. It seems that lethal mutations probably represent complicated metabolic defects that are not reparable with nutritional supplements, or defects in cellular functions that are so fundamental that the cell cannot live without them.

This does not mean, however, that they cannot be studied. There is a class of lethal mutants called *conditional lethals*, which are lethal only under particular environmental conditions. Radiation sensitivity, pH, and osmotic sensitivity have been used, but the most widely used by far are temperature-sensitive mutants, which have wild-type phenotype at their permissive temperature and mutant phenotype at their restrictive temperature. Thus, one incubation temperature allows the temperature-sensitive mutant to survive and grow, whereas a simple change in incubation temperature allows the mutant phenotype to be expressed. No matter how crucial the mutated function might be to the workings of the cell, the mutant can be kept alive at one temperature and studied at another. An example would be mutations that are temperature sensitive with respect to the major DNA polymerase. When the mutation is expressed no DNA can be made, but the mutant grows normally at permissive temperatures, even on minimal medium.

3.8 Isolating Auxotrophic Mutants

The most direct procedure for mutant isolation is to test all survivors from a mutagen treatment individually for their ability to grow on minimal medium. Morphological mutants, usually prototrophic, can be identified by eye, and the auxotrophs fail to grow on minimal medium, but they can be cultivated on complete medium. Identification of the specific nutritional requirement of an auxotroph generally calls for separate culture of the mutants so that each can be tested for growth on a variety of different media in a series of tests, each one narrowing the range of possibilities. In the so-called total isolation procedure, all survivors from a mutagen treatment are cultured individually on complete medium. Inocula are then taken from the cultures for growth tests on minimal medium and minimal medium supplemented with likely growth factors.

Broad categorization can be achieved fairly easily. Minimal medium supplemented with casein hydrolysate supplies all the usual amino acids and supports the growth of amino acid auxotrophs, but no other auxotrophs. Likewise, minimal medium supplemented with hydrolyzed yeast nucleic acid supplies purine and pyrimidines and identifies strains auxotrophic for nucleotide bases, whereas minimal medium supplemented with a synthetic mixture of water-soluble vitamins identifies vitamin auxotrophs. Once a category has been identified, minimal medium supplemented with individual nutrients can be tested. Most auxotrophs respond to a single supplement, although a few with multiple requirements have been found.

An example is the so-called *ad-his* mutant class of *Coprinus cinereus*, which are adenine auxotrophs that show improved growth if histidine is also supplied. These mutants have been shown to lack the enzyme adenylosuccinase, which is responsible for the final step in the synthesis of adenosine 5-phosphate (AMP). The response to histidine is a sparing reaction. That is, because histidine synthesis itself requires AMP, if this requirement is alleviated by exogenous supply of the amino acid, there will be less call on what must be limited amounts of endogenous AMP in a mutant that is still unable to carry out the final step in AMP synthesis even when adenine is supplied.

When isolates have been identified as auxotrophs, specific growth requirements can often be identified using the auxanographic technique. In this, a layer of molten minimal medium agar containing suspensions of spores of the auxotrophic isolate is poured into a Petri dish and allowed to solidify. Small crystals of likely nutrients are then placed on the surface of the agar at well-separated, and marked, points. Following incubation, any of the compounds that support growth are evident from zones of growth around the appropriate crystals. The same technique can be used to assess inhibitory substances.

In 1956 Robin Holliday published a combinatorial method, in which common nutritional requirements are distributed between a small number of pools of nutrients. Each pool contains several different nutrients, and each nutrient occurs in two pools (Table 3.1). The two pools that will support the growth of the mutant indicate the auxotrophic nutrient requirement. For example, with the amino acid pools illustrated in Table 3.1, a mutant that grows on media supplemented with pools 1 and 7 has a requirement for methionine since because is the only amino acid common to both pools. A mutant that grows only on 3 and 6 similarly requires tryptophan. Other nutrients (e.g., vitamins, purines, and pyrimidines) can be incorporated into the pools to suit individual circumstances.

Testing many strains individually in this way can be very labor intensive, so a number of techniques have been devised to reduce the effort involved. The replica plating technique uses a velvet-covered block, sized to fit the surface of the agar medium in a Petri dish. Pressing the velvet to the agar

Table 3.1. Amino acid pools used for discovering auxotrophic requirements.

Pool number	Constituents*
1	Cysteine, methionine, arginine, lysine
2	Leucine, *iso*-leucine, valine
3	Tryptophan, tyrosine, phenylalanine
4	Histidine, threonine, proline, glutamic acid
5	Serine, glycine, alanine, aspartic acid
6	Cysteine, leucine, tryptophan, histidine, serine
7	Methionine, *iso*-leucine, tyrosine, threonine, glycine
8	Arginine, phenylalanine, proline, alanine
9	Lysine, glutamic acid, aspartic acid

* Each pool is made by dissolving 50 mg of each amino acid in 100 ml of distilled water. The pools are then used at the rate of 100 ml/l of minimal medium. A mutant that grows only on media supplemented with pools 1 and 7 has a requirement for methionine because this is the only amino acid common to those two pools. One that grows only on pools 3 and 6 similarly requires tryptophan.

surface then transfers vegetative cells or spores from colonies on a dish of growth medium from one dish to another. The fibers of the velvet pick up the cells and spores, and the pattern of colonies can be "printed" onto other kinds of medium. If all the colonies are able to grow on all kinds of media tested, then the pattern of colonies formed after the imprinted plates have been incubated will be identical on all plates. Any colony on the master (complete medium) plate that is unable to grow on one of the test media, however, will be absent from the pattern of colonies formed on that medium.

The technique was originally designed for use with bacteria and is directly applicable to yeasts and to yeastlike sporidial stages such as those of *Ustilago maydis*, of which 100 or more colonies can be tested on each impression. It can also be used to replicate colonies of filamentous fungi, like *Aspergillus nidulans*, that produce abundant spores while the colonies are still small; the velvet surface just needs to be dampened before use by being pressed onto a sterile agar Petri dish. An adaptation in which the velvet covering of the replicator is replaced by an array of closely spaced pins will replicate fungal colonies whether they produce spores or not. Several hundred fungal colonies can be replicated at a time by incorporating sodium deoxycholate (0.08%) or sorbose (1% with 0.1% normal sugar) into the media to restrict the colony size on the master plates.

Auxotrophs that emerge from a mutagen treatment do not die immediately, so it is feasible to rescue them after a short period of incubation. In practice, a rescue method involves plating mutagen-treated spores on

minimal medium, incubating long enough for prototrophic survivors to form colonies, which are marked, then a layer of medium supplemented with an appropriate metabolite is poured on to the plate and allowed to solidify. Colonies that grow during further incubation after this layering can be isolated, and they usually prove to be auxotrophic for the supplemented metabolite.

The ultimate technique to enhance the isolation of auxotrophic mutants in filamentous fungi is filtration enrichment. This is rather similar to rescue, but it achieves the selective elimination of prototrophic survivors of the mutagenic treatment while retaining most of the auxotrophs. For filtration enrichment, the mutagen-treated spores are suspended in liquid minimal medium and incubated, with aeration, sufficiently for prototrophic spores to germinate and grow slightly. Then the prototrophic growth is removed by filtration through some kind of sterile filter in which the hyphae of germinated spores are held but ungerminated spores are allowed to pass through. Incubation and filtration cycles are repeated until essentially all prototrophic spores have been removed. The remaining suspension can then be spread on complete medium for total isolation or onto minimal medium supplemented with metabolites of interest for selective rescue of particular auxotrophs.

The nature of the filter, the length of the incubation period and the number of incubation–filtration cycles are the crucial variables that have to be established by pilot experiments before filtration enrichment can be applied efficiently to an organism. The filter must trap germinated spores efficiently without trapping ungerminated ones and this is a compromise between the amount of growth made (and its growth pattern, degree of branching, etc.) during incubation of the suspension and the filter material itself. Several medical products like muslin, "cotton wool," and medical gauze achieve the necessary balance between retention of hyphae and wash-through of spores, the choice depending on the hyphal growth pattern of the fungus concerned. Make sure you use wettable varieties, though; waterproofed products do not work!

The incubation period should be long enough to produce prototrophic growth only just visible to the eye, and no longer. It obviously depends on the rate of germination and growth of the fungus concerned, but is likely to be in the 1–4 day range. A fully optimized process can result in more than 10% of the colonies that are finally plated being of the desired auxotrophic types. Possible faults, however, are that filtration enrichment can yield mutants which grow slowly on any medium, and selects against certain types of auxotrophs. Some auxotrophs can grow on the minute traces of metabolites (like some vitamins) that can appear in the suspension through breakdown of dead spores, and such auxotrophs will be filtered off. Other auxotrophs (e.g., inositol-requiring mutants) often die rapidly in minimal medium and fail to survive the incubation periods of the filtration enrichment process.

Use has been made of the relative survival in minimal medium in another mutant selection process. Mutants with two nutritional deficiencies sometimes survive longer in minimal medium than either single mutant, and this is the basis for an effective method to obtain auxotrophs. Spores of a biotin- or inositol-requiring auxotroph can be exposed to mutagen to 5% survival, then spread on minimal medium and covered with a further layer of minimal medium. During incubation of the plates at least 99% of the original, single mutant, spores die. A further layer of complete medium is then poured onto the plate, and up to 60% of the colonies that grow after this final layer is added will be (new) double auxotrophs. The physiological basis for this method seems to be that the original vitamin-requiring auxotroph is subject to unbalanced growth, in which the spores germinate but are unable to sustain growth because of the vitamin deficiency; thus, they effectively commit suicide. Germination and suicide is prevented by the second auxotrophic mutation.

Another approach that achieves selective elimination of prototrophs is to use antibiotics or other drugs to kill by inducing unbalanced growth. This has its origins in a method for auxotroph isolation in the bacterium *Escherichia coli* that uses the fact that penicillin kills growing bacteria because it prevents wall formation. As a result, in the presence of penicillin on minimal medium prototrophs die (they attempt to grow but are killed), whereas auxotrophs survive because they can't grow in the absence of the nutrient they require. Fungi are not sensitive to penicillin, but there are some other antibiotics (e.g., amphotericin B, endomycin, nystatin, N-glycosyl polifungin, neotropsin and 2-deoxyglucose) that allow the same approach to be applied to fungi, and it can increase the proportion of auxotrophs in the mutagenised population by about 50 times. Even radiation has been used to eliminate prototrophs selectively. Prototrophs of *Saccharomyces*, *Neurospora*, and *Aspergillus* grown on minimal medium will incorporate RNA or protein precursors labeled with tritium (^{3}H). If stored at low temperature, the cells are killed by the β-particles that result from tritium decay. On the other hand, auxotrophic (or temperature-sensitive) mutants that were not active during exposure to tritiated metabolites tend to survive, so the proportion of mutants is enriched.

Over the years a number of clever techniques have been used to aid the selection of particular mutants, so let's look at a number of these "tricks of the trade." Brewing and wine yeasts with higher alcohol tolerance have been selected *in vitro* by selecting for yeast strains able to grow on medium containing high alcohol concentrations. Higher alcohol tolerance in the yeast gives higher alcohol concentrations in the product, and merrier consumers (well, actually, more profits for the brewer). It is possible to select for high yields of excreted products (e.g., extracellular enzymes, amino acids, or antibiotics) using an overlay method. For enzymes, the overlay contains the enzyme substrate. After incubating to produce some growth of the putative mutants on an appropriate medium the overlay is poured over the

plate and breakdown of the substrate in the overlay is detected after some further incubation.

Enzyme activity can be detected as visible clearing zones in translucent overlays that contain poorly soluble substrates, like starch, cellulose, or casein; or by adding further overlays of chemicals that stain the substrate or detect breakdown products. Proteinases, cellulases, amylases, lipases, pectinases, and tannin-degrading enzymes are all good candidates for this type of detection. Mutants unable to produce any one of these enzymes may fail to form a clearing zone on the relevant substrate; however, mutants with defects in the general pathway responsible for export of extracellular proteins will fail to form clearing zones on several different substrates. The overlay might include other organisms. For example, a mutant fungus that overproduces and excretes an amino acid might be detected by an overlay containing bacteria that are auxotrophic for that amino acid. Antibiotic producers can similarly be detected by overlays containing the organisms that are the target of the desired antibiotic.

3.9 Resistance Mutations

The forward mutations discussed so far result in some sort of functional defect that is expressed under normal growth conditions but can be repaired by altering the conditions in some way. *Resistance mutations* are forward mutations in organisms that have gained the ability to grow under conditions not suitable for growth of the wild type. The selective advantage these resistance mutants have under these conditions means that spores can be placed at high densities on medium containing the compound (or incubated under the conditions), which inhibits wild-type growth. Drug resistance is an obvious example for which mutants have been selected in many fungi.

Many chemical analogues of normal metabolites are inhibitory because they are sufficiently similar to the normal metabolite to be incorporated in metabolism, but sufficiently different to cause a critical breakdown in metabolism. Resistance to antibiotics or inhibitors can be due to several types of mutations: in genes coding for, or regulating, the proteins or metabolic steps sensitive to the inhibitor in the wild type; in other areas of metabolism that enable the inhibitor to be detoxified; or in uptake mechanisms so that antibiotics/inhibitors no longer enter the cell in sufficient quantity to cause inhibition. Some chemicals become inhibitors after metabolism "activates" them in some way; in these cases, mutations in genes involved in the activating reactions will be another way of conferring resistance. Pathogenic fungi are under continuous attack by their hosts (not to mention humans, who try to destroy them with pesticides). This also imposes selection pressure, which can result in the emergence of strains that have gained virulence.

3.10 Reverse Mutation: From Auxotroph to Prototroph

Resistance mutants can be selected because the mutant phenotype allows them to grow on medium on which the parental strain fails to grow. The same logic applies to reverse mutations from auxotrophy to prototrophy. These mutations can be detected readily simply by plating spores of the auxotroph at high density on minimal medium. Reverse mutation (or reversion) to prototrophy is readily selected and all aspects can be quantified (e.g., mutagen exposure, if any, number of spores plated on minimal medium, and number of prototrophs obtained. There are two potential complications: the Grigg effect and the occurrence of suppressors.

Reconstruction experiments, carried out by G.W. Grigg with *Neurospora* conidia, in which known numbers of prototrophic spores were added to the auxotroph suspensions showed that germination of prototrophs could be adversely affected by the presence of large numbers of auxotrophic spores. It seems that although they were unable to grow, auxotrophic spores respired sufficiently to cause competitive inhibition of the prototrophs on the sorbose-containing medium used with *Neurospora*. Adapting the medium and adjusting the spore density can avoid the problem.

Reversion rates are allele specific and very variable because reversion to prototrophy can result either from true back-mutation at the original mutant base pair or to suppressor mutations either within the same gene or in another gene. Suppression due to a second forward mutation in the same gene (called *intragenic suppression*) is generally caused by a second mutation re-establishing the correct reading frame in a strain carrying a frame-shift mutation. If the initial frame-shift mutation is caused by deletion of a single base, then the insertion of one base a few base pairs downstream can result in intragenic suppression. The sequence between the two mutations will still be translated into the wrong amino acids, so the effectiveness of the intragenic suppressor mutation will depend on how well the gene product can function despite the presence of those abnormal amino acids.

Reversion due to suppression of the mutant phenotype by a second mutation in a different gene (called *intergenic suppression*) may occur because the suppressor mutation opens up an alternative pathway supplying whatever metabolite was deficient in the original auxotroph. This is also called *metabolic suppression*, and such suppressors are gene-specific rather than allele-specific. An allele-specific metabolic suppressor can act by changing intracellular conditions so that a mutant enzyme regains activity. For example, a mutant of *Neurospora crassa* that produced a mutant tryptophan synthase sensitive to inhibition by zinc was suppressed by a second mutation that reduced zinc uptake. Most allele-specific suppression, though, results from mutation in a transfer RNA that corrects the coding defect of the original auxotroph by altered reading of the genetic code at translation (but leaves the original mutation unchanged).

With appropriate allowance for these complications, mutation frequencies can be measured and related to treatments applied to the auxotrophic spores. A good deal of what we know about mutation in eukaryotes was learned with fungi using this approach to study mutation kinetics. It is generally found that the back-mutation frequency (i.e., the reversion frequency excluding suppressors from consideration) is lower than the frequency of forward mutation. This is because mutations at many different sites can abolish gene function, but very specific change is required to restore function to a mutant gene. The usual expectation is that forward mutation frequencies tend to be 10–100 times higher than reverse mutation frequencies.

3.11 Molecular Variants

The mutants (or variants) we have discussed so far all contain mutations of functional genes. Their selection is based on the formation/expression (or lack thereof) of a gene product; however, there are several features of DNA structure, which do not necessarily involve functional genes, that serve as valuable DNA markers in genetic analyses and that can be used to select for mutants/variants in the population (and see Chapter 7).

The first DNA markers of this sort were RFLPs (Restriction Fragment Length Polymorphisms). Restriction fragments are the pieces of DNA produced by digestion with a restriction enzyme. *Restriction enzymes* are endonucleases found in various bacteria, which cleave DNA at specific sequences in the double helix. Their purpose in bacteria is to provide a defense mechanism against such extraneous DNA as viruses and plasmids, but the specificity with which they cleave DNA makes them extremely useful in molecular biology. The most widely applied restriction enzymes recognize very specific sequences of 4, 5, 6, or 7 nucleotides that are palindromic. For example, a very commonly used restriction endonuclease is called *Eco*RI and it cuts double-stranded DNA at the sequence:

5′-G↓AATTC-3′

3′-CTTAA↓G-5′

by cleaving phosphodiester bonds between the guanine and adenine residues (indicated with ↓). The importance of restriction enzymes lies in the fact that samples of any one sort of DNA are always cut at the same sites; thus, the same set of fragments will always be produced. The fragments can be characterized, essentially by molecular mass, from their pattern of migration by electrophoresis through a gel. That migration pattern can be visualized using fluorescent or radioactive probes, which bind to DNA at specific sites; the pattern is effectively the restriction phenotype of that sort of DNA. DNA from the same strain will always yield the same restriction phenotype, but DNA sequences can vary between different isolates of a

species. Because restriction enzymes cleave DNA at specific base sequence sites, when the location of these sites varies between different genotypes the DNA fragments produced will vary. Differences between strains lead to differing DNA fragment patterns, or different restriction phenotypes. In this way the genotypes of the strains can be distinguished on the basis of the patterns of bands on an electrophoretogram, and each difference between genotypes is termed an RFLP (see Section 7.3 for more detail).

As molecular markers, RFLPs can be used in a very wide variety of genetic analyses. They can be used to construct genetic linkage maps, and particular RFLPs can be associated with, and used as a marker for, other traits of interest. Another application of RFLPs is in DNA fingerprinting, where the particular pattern of DNA fragments is used as an identifier for a particular individual, isolate, or strain (see section 7.7). They have several advantages over mutations scored as changes in gene function. RFLPs do not show dominance–recessive relationships (RFLP alleles are said to be codominant), and there is only one phenotype (pattern of bands on an electrophoretogram) so there is no complication with pleiotropic effects; however, they do exhibit Mendelian segregation.

The frequency of RFLPs should indicate the extent of variation between strains. A lack of DNA polymorphism could reflect very close genetic relationships of isolates; however, there is more opportunity for polymorphism at the DNA level than in any functional product of a gene. This is because functional mutations are only detectable if they occur in that part of the DNA that is expressed. The restriction phenotype covers all of the DNA, whether it is transcribed or not. There are a number of other potentially useful molecular markers that are conceptually similar to RFLPs, but which differ in the nature of their sequence; these will be discussed in Chapter 7.

Classic genetic analysis starts with mutants in which a phenotype has become altered, and the objective is identification of the genetic control of that phenotype. With the advance of molecular methods of genome analysis the strategy has changed. Large-scale sequencing reveals genes as sequences of no known function and the challenge then is to assign function. This still depends on mutating the gene to establish what function is consequently lost, but random mutation with mutagens is of no use when the need is to inactivate a specific gene sequence. The technique that has been developed is called *site-directed mutagenesis*. It works by disrupting the target gene with an unrelated segment of DNA by recombination between the resident chromosome and a piece of transforming DNA that has enough homology with the target gene to locate it on the chromosome.

A technique employed in *Saccharomyces cerevisiae* uses a deletion cassette that includes an antibiotic resistance gene with segments of DNA at either end of it that have the same sequence as that of the target for inactivation. Yeast cells are transformed with the deletion cassette, and transformed yeasts in which the replacement has taken place successfully are then selected on medium containing antibiotic. Transformed cells are antibi-

otic resistant, thanks to the resistance gene included in the cassette, but untransformed cells are sensitive to the antibiotic. The transformant phenotypes can subsequently be studied to determine the function of the target gene. There is an analogous procedure for use with mice; it uses exactly the same recombination strategy, but it is applied to embryonic stem cells, which are then injected into mouse embryos, eventually producing gene knockout mice. Site-directed mutagenesis can be used to study gene function in great detail, but it requires prior knowledge of the DNA sequence of interest as well as reliable sequence manipulation and transformation protocols, which are available for a relatively small number of fungi.

Gene disruption using transformation and homologous recombination is increasingly used to obtain mutants of a specific sort. A development of the technique described in the previous paragraph is known as restriction enzyme mediated insertion (REMI). In this approach, fungi are transformed with a DNA construct, but restriction enzymes are included in the transformation mixture. Addition of low concentrations of restriction enzymes to transformation mixtures results in earlier appearance and improved yield of transformants. There are a number of problems associated with disruption of fungal genes. The optimal restriction enzyme identity and concentration need to be established in pilot experiments because transformation rates decline at high enzyme concentrations, and the enzymes have cytotoxic and other damaging effects.

Other difficulties worth remembering are that disruption of a specific DNA sequence may not be recognized as a distinct phenotype because other (undisrupted) genes replace the lost function; that is, many gene functions in eukaryotes are redundant. In addition, a disruption phenotype may not be detected simply because not enough is known about the particular conditions and/or the particular stage in development at which the sequence of interest is expressed. Nevertheless, the approach is being used successfully to generate mutants in particular aspects of the life style of the experimental organism (e.g., pathogenicity, morphogenesis, and development). A new method uses *impala*, an active transposable element of *Fusarium oxysporum*, as a tool to attack particular genes. Engineered *impala* copies have been introduced with good efficiency into *Fusarium moniliforme*, *Aspergillus nidulans*, *Aspergillus fumigatus*, and *Magnaporthe grisea*, which indicates that *impala* can be used to tag and inactivate genes from a wide range of filamentous fungi.

Publications and Websites Worth a Visit

Authoritative information about numerous carcinogens and mutagens can be found in the Ninth Report on Carcinogens (as revised January 2001), which is produced regularly by the U.S. Department of Health and Human Services in the Public Health Service National Toxicology Program. The current report can be accessed and downloaded from http://ehis.niehs.nih.gov/roc/.

Bos, C.J. & Stadler, D. (1996). Mutation. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 13–42. Marcel Dekker, Inc.: New York.

Cummings, W.J., Celerin, M., Crodian, J., Brunick, L.K. & Zolan, M.E. (1999). Insertional mutagenesis in *Coprinus cinereus*: use of a dominant selectable marker to generate tagged, sporulation-defective mutants. *Current Genetics* **36**, 371–382.

The Environmental Mutagen Society (EMS) is the primary scientific society fostering research on the basic mechanisms of mutagenesis as well as on the application of this knowledge in the field of genetic toxicology. The EMS website is at http://www.ems-us.org.

Granado, J.D., Kertesz-Chaloupková, K., Aebi, M. & Kües, U. (1997). Restriction enzyme-mediated DNA integration in *Coprinus cinereus*. *Molecular and General Genetics* **256**, 28–36.

Kahmann, R. & Basse, C. (1999). REMI (restriction enzyme mediated integration) and its impact on the isolation of pathogenicity genes in fungi attacking plants. *European Journal of Plant Pathology* **105**, 221–229.

Mutation Research is a research journal that published a special issue entitled *Fundamental and Molecular Mechanisms of Mutagenesis* in November 2000 (*Mutation Research* Volume **455**, Issues 1–2, pp. 1–215). This is expected to be the first volume in an Elsevier series on *Protocols in Mutagenesis*, and contains 13 papers providing bench-level details for studies using different mutagenesis assay systems.

Pukkila, P.J. (1996). Production and analysis of meiotic mutants in *Coprinus cinereus*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 363–369. Marcel Dekker, Inc.: New York.

Radford, A., Bruchez, J.J.P., Taleb, F. & Stone, P.J. (1996). Mutation in *Neurospora crassa*: from X-rays to RIP. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 281–294. Marcel Dekker, Inc.: New York.

Risk Management Internet Services provides a family of resources for risk management at http://rmis.com. Features on offer include a listing of databases dealing with chemicals at http://rmis.com/sites/cheatab.htm.

Sanchez, O., Navarro, R.E. & Aguirre, J. (1998). Increased transformation frequency and tagging of developmental genes in *Aspergillus nidulans* by restriction enzyme-mediated integration (REMI). *Molecular and General Genetics* **258**, 89–94.

Sweigard, J.A., Caroll, A.M., Farrall, L., Chumley, F.G. & Valent, B. (1998). *Magnaporthe grisea* pathogenicity genes obtained through Insertional mutagenesis. *Molecular Plant-Microbe Interactions* **11**, 404–412.

TOXNET is a website giving access to a cluster of databases on toxicology, hazardous chemicals, and related areas that provide factual information on toxicity and other hazards of chemicals. The website is at http://toxnet.nlm.nih.gov/.

Wright, B.E. (2000). A biochemical mechanism for nonrandom mutations and evolution. *Journal of Bacteriology* **182**, 2993–3001.

Historical Publications Worth Knowing About

Evans, H.J. & Lloyd, D.C. (1978). *Mutagen-Induced Chromosome Damage in Man*. Edinburgh University Press: Edinburgh.

Holliday, R. (1956). A new method for the identification of biochemical mutants of micro-organisms. *Nature* **178**, 987.
Venitt, S. & Parr, J.M. (1984). *Mutagenicity Testing: A Practical Approach*. IRL Press: Oxford, UK.
Wood, R.D. & Sedgwick, S.G. (1986). Molecular aspects of mutagenesis. *Mutagenesis* **1**, 399–405.

CHAPTER 4

Segregation Genetics: The Indirect Approach

Revision Concepts

- Complementation tests determine whether similar phenotypes result from mutation in the same gene (functionally identical) or different genes (functionally different).
- Complementation groups indicate which mutants are alleles of the same gene locus; noncomplementation = allelism = no growth on minimal medium by confrontations between auxotrophic mutants.
- To verify complementation groups, precursors are fed to auxotrophs to determine the point in the pathway that is blocked by mutation; another strategy is to examine mutant mycelium for accumulation of precursors.
- Allelic complementation results when the respective polypeptides of two allelic mutants "repair" each other when they bind together, partially to recover function.
- Complementation maps showing the relationships between complementing alleles probably reflect the molecular interactions between the gene products.
- Confrontations between resistance mutants will not grow on medium containing inhibitor if the mutant alleles complement one another, but will grow if they do not complement.
- Difficulties with the complementation method arise when the heterokaryons under study are unstable or when the nuclear ratios are not equal.
- Complementation tests cannot be done with dominant mutations, with molecular variants that are not functional phenotypes, and with genes that show epistasis.
- In ascomycetes karyogamy occurs in the ascus mother cell, meiosis occurs in asci, and is usually followed by a mitotic division so that eight uninucleate, haploid ascospores are typically formed.
- In basidiomycetes karyogamy and meiosis take place in the basidium, and basidiospores (usually four) are produced externally.

- Karyogamy and meiosis in heterokaryotic fungi occur as they do in other eukaryotes following Mendelian laws of segregation, independent assortment, linkage, and crossing over.
- Landmarks in the meiocyte developmental pathway are commitment to recombination, commitment to the first and second meiotic divisions, commitment to spore formation, and commitment to spore maturation.
- Premeiotic DNA replication, homologous chromosome pairing, and formation of the synaptonemal complex are independent events.
- Unlike other eukaryotes, fungi carry out meiosis with the nuclear membrane intact in prophase I and the leptotene stage may be brief or absent in fungi.
- If chromosomes are segregating independently, then the four genotypes of the haploid spores produced will be formed in equal frequency.
- Statistical tests (i.e., the χ^2 significance test) determine whether numerical deviations of the predicted progeny ratios are acceptable.
- The χ^2 test indicates the probability that deviations between observation and expectation are not significant (due to chance variation) or significant (due to a cause other than chance).
- High probabilities indicate good agreement between observed and expected results.
- The χ^2 distribution can be applied to contingency tables to test the homogeneity of samples from different experiments and determine whether they share the same distribution.
- Recombination frequency is a measure of the distance between two genes and can be used to construct a map.

4.1 Complementing Mutants

To study the genetic basis of any phenomenon, a collection of genetic variants is an essential resource. Whether these are collected from the natural population or obtained from induced mutation hunts, the chances are high that the final collection will contain several variants with the same phenotype. The immediate question this raises is whether they have the same phenotype because they carry mutations in the same gene, or whether they are mutated in different genes that all happen to influence the phenotype being scored.

To answer this question, we have to determine whether the mutations are functionally different or functionally identical. This can be done with recessive mutations if you can bring the two mutants together into the same cell. If the mutations have occurred in the same gene (i.e., if they are alleles) the cell will be homozygous for the mutation and will consequently express the mutant phenotype. On the other hand, if the mutations have occurred in different genes, when the mutant nuclei are brought into the same cell they each bring a wild-type version of the gene that is mutant in the other. Each

mutation, therefore, will be heterozygous and the cell will express the wild-type phenotype. This is the standard functional test for allelism. It is called a *complementation test* and depends on the fact that genes are expressed through the production of mRNA and protein products into the cytoplasm. If both copies of a gene are mutated there is no way to produce active gene product and the mutant phenotype results, but even a single normal copy of a gene is sufficient to provide normal function so the double heterozygote has the wild-type phenotype.

It is usually necessary in animals and plants to complete a genetic cross and produce a diploid to carry out a complementation test. Fungi can make heterokaryons, though, and because complementation depends on creation of gene products into the cytoplasm, the heterokaryon is a suitable tool. To explain how the complementation procedure is used and how it contributes to further studies, we will describe some adenine-requiring mutants of *Coprinus cinereus* as an example.

4.2 Adenine Auxotrophs of *Coprinus*

Approximately 70 independently induced mutants with a requirement for adenine have been isolated from wild strains of *C. cinereus*, but we will use only a few of these to make our point. The mutants have been isolated at different times by different people using a number of mutagenic treatments, so their origin is no guide to relationships. The study was designed to allocate the mutants to their gene loci so that these could be correlated with steps in the purine biosynthetic pathway.

The test for functional allelism between different auxotrophic mutants in *C. cinereus* is for growth of their mutual dikaryon on minimal medium. The two mutants are unable to grow on minimal medium as monokaryons: they are auxotrophs and the medium must be supplemented with adenine before they can grow. Vigorous growth of their joint dikaryon on minimal medium shows that the two mutations must be in different gene loci (i.e., the mutants complement one another); they are not allelic. Dikaryons can be made on a complete medium and small inocula subsequently transferred to minimal medium for the growth test. Because the mutants carry the mating-type specificity of the wild type used in the initial mutagen treatment, the first step in the procedure is to cross each mutant with a compatible wild type to produce mating-type recombinants. This enables isolation of a panel of mutant strains with compatible mating types so that the dikaryons can be constructed in all possible combinations, giving rise to a complementation table or complementation matrix like Fig. 4.1.

You initially have no way of sorting the mutants under test except for whatever isolation code or culture collection stock number is use to identify them, which is how the strains are arranged in the left-hand panel of Fig. 4.1. The results are entered in this table with closed circle, ●,

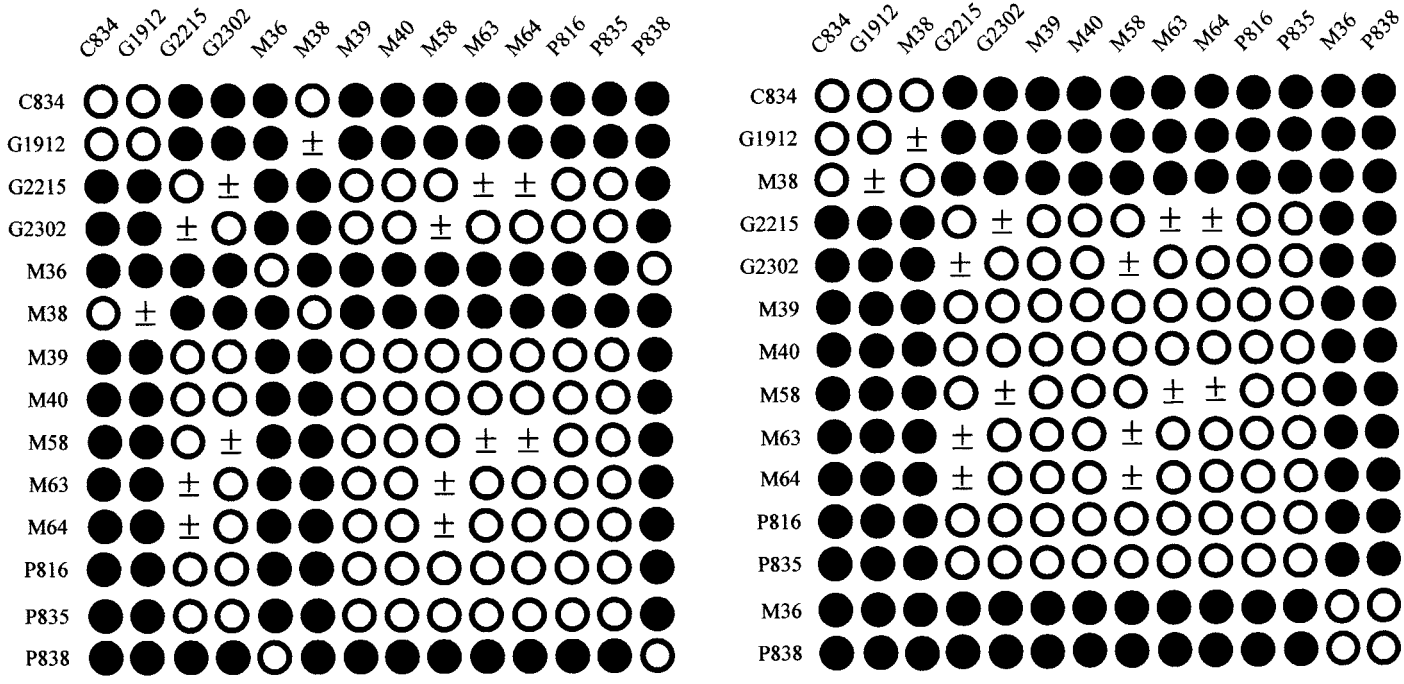


Fig. 4.1. Complementation matrix of adenine auxotrophs of *Coprinus cinereus*. The data in the panel on the right has been sorted to bring noncomplementing results into common blocks.

symbolizing vigorous growth of the dikaryon on minimal medium (= complementation between the mutants), and an open circle symbolizing no growth (= noncomplementation). You will notice that there are a few cells in the matrix in which the plus/minus symbol, ±, is entered, which indicates that a small amount of mycelial growth was detected, much less than could be described as "vigorous" like the wild type, but still not wholly mutant phenotype either. These tests have to be repeated to verify that the medium was not contaminated by the nutrient the mutants require, and any hyphae that grow on the minimal medium must be checked to be sure they are dikaryotic (it's quite possible for a mutant to back-mutate, or revert, to wild type, and the transfer to minimal medium could have selected one of these rare revertants). If these possibilities can be excluded, then the explanation is that two allelic mutants are partially correcting for each other's mutation. This is a phenomenon called *allelic complementation* to which we will return at the end of this section.

For the moment, note that the complementation matrix in Fig. 4.1 includes pure homozygous dikaryons, such as [C834 + C834], [G2302 + G2302], and the like. These are essential negative controls that show how the mutant behaves in the test regime and, more importantly, in the dikaryotic condition. The matrix also features a complete set of reciprocal crosses, such as [C834 + G2302] as well as [G2302 + C834]. These are in-built replicate tests of the mutant alleles, but remember that the strains must have compatible mating types, so these are not identical dikaryons; rather, they may be something like [C834, A_1B_1 + G2302, A_2B_2] and [G2302, A_1B_1 + C834, A_2B_2] (i.e., true reciprocal crosses rather than simple replicates). Providing the top-right triangle of the matrix is a mirror image of the bottom-left triangle, these reciprocals verify that the mutant phenotype has been stably transmitted through the genetic cross used to make the mating-type recombinants. When the tests have been completed (and consistent results obtained over several repeated trials), interpreting the matrix involves rearranging the data it contains so that all the negative results are brought together into blocks, which are called *complementation groups*. This has been done in the right-hand panel of Fig. 4.1.

After sorting the information, it's evident that the data fall into three complementation groups, as indicated by the blocks of open circles in Fig. 4.1. Members of a complementation group are alleles of the same gene locus. For example, isolates C834, G1912, and M38 are independent mutations in the same gene. Because the product of this gene contributes to synthesis of adenine (the product is presumably an enzyme in the metabolic pathway), mutants in the gene are adenine auxotrophs, and the convention is to assign it the gene symbol *ade*. As this is the first of three adenine loci we will be naming, it will be called *ade*-1. Likewise, the complementation group made up of isolates G2215, G2302, M39, M40, M58, M63, M64, P816, and P835 are independent mutations in a functionally different gene, which can be called *ade*-2. The next two genes in the sequence were named using

strains and data not represented in this example, so the complementation group comprising isolates M36 and P838 is called *ade*-5.

This analysis revealed three functionally distinct complementation groups; *ade*-1, *ade*-2, and *ade*-5. That these really do represent different genes whose products control different steps in the biosynthesis of adenine was demonstrated with further biochemical analyses. With many auxotrophs this can be done by feeding precursors in the pathway on the expectation that an auxotroph will grow if supplied anything in the pathway that occurs after the step that is blocked by the auxotrophic mutation, but will not be able to grow on medium supplemented with a precursor that occurs in the pathway before the block. The majority of the intermediates in purine biosynthesis are unfortunately rather exotic compounds; they are not readily available, and are not readily taken up by the cell. Another useful strategy in such a case, however, is to examine the mutant mycelium for accumulation of precursors.

The purine biosynthetic pathway is a good candidate for this approach because the aminoimidazoles that are likely to accumulate are relatively easy to isolate, separate with paper chromatography and characterize with spectrophotometry. The major accumulant of *C. cinereus ade*-1 was identified as aminoimidazole ribotide (AIR), which indicates an inability to carry out the carboxylation of AIR. Alleles of *ade*-2 did not accumulate imidazoles and must be blocked in a step that occurs before imidazole ring formation. A compound called 5-amino-4-imidazole-*N*-succinocarboxamide ribotide (SAICAR) accumulated in the mycelium of *ade*-5 alleles. When coupled with the fact that *ade*-5 alleles could not use hypoxanthine for growth whereas all other loci did, this places the *ade*-5 block in a reaction specific to adenosine-5-phosphate (AMP) synthesis and must lack activity of the bifunctional enzyme adenylosuccinase. In summary, the three functionally distinct complementation groups, *ade*-1, *ade*-2, and *ade*-5 can be assigned to three distinct enzymes in the purine biosynthetic pathway (Fig. 4.2), which demonstrates that the (relatively simple) complementation test very effectively identifies discrete gene functions.

The three complementation groups indicated in Fig. 4.1 can be established readily despite the potentially confusing instances of allelic complementation (the ± symbols in the original matrix). Allelic complementation is usually rare. On average only about 10 percent of all tests will give rise to allelic complementation, so the members of a complementation group are "interlocked" with each other by the great majority of definitive tests. This is illustrated in Fig. 4.1 by the three alleles of *ade*-1. G1912 and M38 show allelic complementation, but any thought that this might be full functional complementation (indicating that they are alleles of different genes) is removed by the fact that both of these isolates fail to complement isolate C834. These data can be represented in a complementation map. The conventions for drawing such a map are that (1) an allele is represented by a horizontal line, (2) lines that correspond to noncomplementing alleles

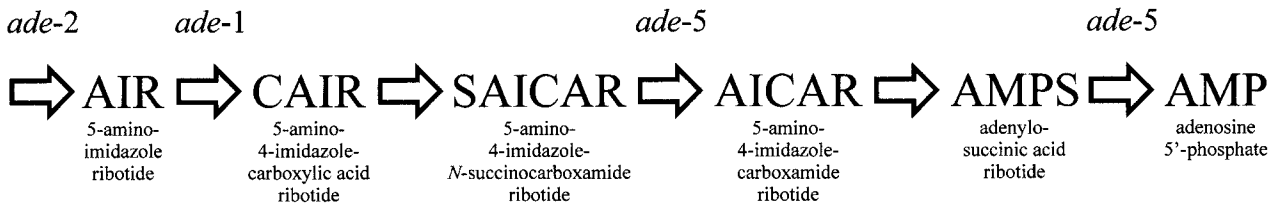


Fig. 4.2. The purine biosynthetic pathway, showing the involvement of some adenine auxotrophs of *Coprinus cinereus*.

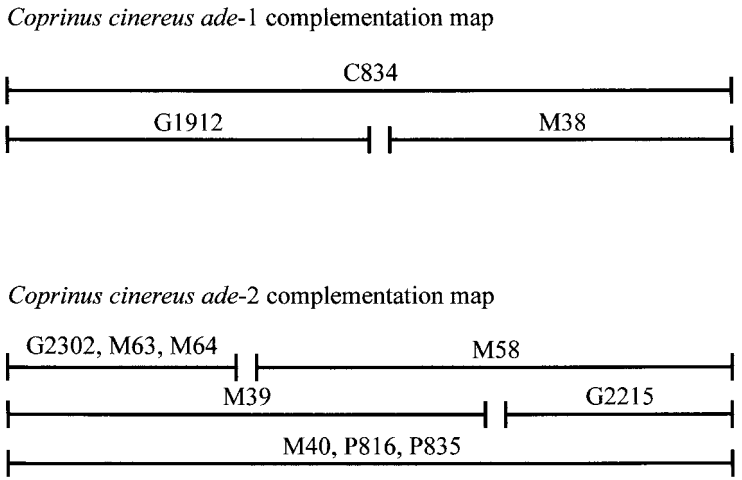


Fig. 4.3. Complementation maps corresponding to the complementation data in Fig. 4.1.

overlap, and (3) lines corresponding to complementing alleles do not overlap.

Applying these conventions to *ade*-1 gives the map shown in the upper half of Fig. 4.3. Because there are only three alleles to be represented, this is obviously a very simple map. The more complementing alleles there are, the more complex the map must be. The complementation map corresponding to the complementation data for the *ade*-2 locus of Fig. 4.1 is shown in the lower half of Fig. 4.3.

Allelic complementation is interpreted as being the result of interactions at the polypeptide level. Many, perhaps the majority, of the gene functions are carried out by proteins that consist of several subunits (i.e., they are multimers), where the subunits (or monomers) are the polypeptides specified by the gene sequence. When a gene is mutated to a defective allele, all polypeptides encoded by that gene in the cytoplasm are defective, the final multimer is also defective, and the gene function is lacking. If two mutant alleles are brought together into the same cytoplasm, however, it is conceivable that conformational defects in the monomers might be partially corrected as they bind together in hybrid multimers. This could restore some function and, consequently, result in allelic complementation. In a sense, therefore, allelic complementation is a means of studying interactions between polypeptides.

In the 1960s through to the 1980s, this was the *only* way of investigating such conformational interactions, and some enormously detailed analyses were done. The resultant allelic complementation maps were extremely complicated and so difficult to interpret that little was achieved. With the

benefit of hindsight, we can appreciate that complementation maps are the two-dimensional projections of three-dimensional interactions between polypeptides (indeed, even four-dimensional interactions because complementation tests are growth tests they potentially reflect interactions over the fourth dimension of time). As a result, allelic complementation maps are bound to be difficult to interpret. Hindsight is a great aid to understanding!

4.3 Functional Allelism

Complementation tests for functional allelism can be done with any recessive variant. All that is required is a clear distinction between wild type and mutant phenotypes. With auxotrophic mutants, like the adenine-requiring mutants of *Coprinus* described earlier, we have seen that this distinction results in the complementing pairing forming a heterokaryon that will grow readily on minimal medium, whereas the noncomplementing heterokaryon will not grow readily on minimal medium. With resistance mutants the distinction is reversed. For example, say you are studying mutants selected for resistance to a growth inhibitor like the sugar analogue 2-deoxy-D-glucose. In this case, the complementing heterokaryon will not grow readily on medium containing 2-deoxy-D-glucose, whereas noncomplementing heterokaryons will grow readily on medium containing the inhibitor. The rule is: complementing heterokaryons show the wild-type phenotype; noncomplementing heterokaryons have the mutant phenotype.

The approach can also be used with morphological mutants, providing, as usual, that the mutations are recessive and that the distorted morphology does not interfere with heterokaryon formation. Some colonial mutants of *Fusarium venenatum* provide an interesting example. Colonial mutants are more highly branched than the wild type and have greatly reduced radial growth rates. As a result, they form much more compact colonies than usual. The mycelium of *F. venenatum* is grown to make the mycoprotein Quorn™ product. The marketing of Quorn™ emphasizes the product to be a low-fat, low-calorie, cholesterol-free, healthy alternative to meat. The nutritional aspects of these claims obviously depend on the fungal origin of the material, but the sparsely branched filamentous hyphal structure of the normal *Fusarium* is crucial to the creation of the meatlike texture of Quorn™ product. Highly branched colonial mutants do not have the filamentous character that is essential for the product.

Colonial mutants unfortunately have some (still unknown) selective advantage over the parental strains in the continuous fermenters used for production. When a colonial mutant does arise, its population increases in the fermenter, reaching a level at which the harvested mycelium no longer has the structure needed for the Quorn™ product. At that point the fermenter has to be closed down and the whole process reinitiated with a fresh culture of wild-type mycelium. Such an interruption in production is

enormously costly. Understanding how colonial mutants arise and what selective advantage they have would allow significant progress to be made toward controlling this problem. We will discuss this in more detail in Section 9.11.

The relevance here is that the first step is to identify how many functional genes might be involved. This has been done using complementation analysis. Colonial mutants were isolated as they occurred, until a collection of 20 strains was assembled and these were used to make heterokaryons. Morphology and growth rate of the heterokaryons were measured to compare with wild type; this was not always easy because the heterokaryons of *F. venenatum* were generally unstable. Another difficulty with many fungi is that the ratio of the two types of nuclei in the heterokaryon is not always equal and is rarely known. The analysis of adenine-auxotrophs of *Coprinus* described earlier did not suffer these problems because the organism forms very stable dikaryotic heterokaryons, which regularly have one nucleus of each type in each hyphal cell. In many other fungi, including some basidiomycetes, nuclear ratios are not so closely regulated, and this has to be taken into account by devising quantitative tests with adequate controls. Suitable tests were devised for *F. venenatum* colonial mutants in which nuclear ratios were monitored by examining morphologies of colonies formed by spores that arise on the heterokaryons. It was found that the 20 mutant strains fell into three complementation groups, which are presumed to correspond to three genes.

Complementation tests cannot be used with dominant mutations because homozygotes (i.e., noncomplementing pairings) and heterozygotes (i.e., complementing pairings) have the same (mutant) phenotype. They also cannot be used with molecular variants (see Chapters 8 and 9) where the phenotype is some feature of the DNA (e.g., migration on a gel or the base sequence, which is a physical rather than a functional phenotype). Epistasis is another potential complication. *Epistasis* is the situation in which one gene, the epistatic gene, masks the expression of a second gene, the hypostatic gene. In effect, the epistatic gene is "dominant" to the hypostatic one, and this effect, like true allelic dominance, can influence the phenotype produced when the genes are brought together. For example, if you are studying morphology of a fruit body, any gene mutant that influences the morphology of the fruit body will be hypostatic to a mutant that prevents fruit body formation. Any culture carrying the latter gene will not make fruit bodies no matter what other genes it might contain that determine fruit body morphology.

It is possible, however, to use epistasis to establish the order in which genes operate in a common pathway. For example, if you make double mutants carrying genes *x* and *y*, with the two genes affecting different aspects of the same gross phenotypic characteristic, and where they all have *only* the phenotype of gene *y*, then *y* is epistatic to *x*. Further, it is a reasonable interpretation that the *y* function must be carried out before the

x function can be expressed. Thus, the pathway must run from *y* to *x*. In Section 10.12 we will describe the use of epistasis to deduce the order in which conidial mutations function during conidiophore morphogenesis of *Aspergillus*. The approach has also been used to establish the order of events in the pathway of protein secretion in yeast.

4.4 Gene Segregation Depends on the Behavior of Chromosomes During Nuclear Division

In mitosis, the haploid nucleus of a vegetative hyphal cell divides to produce two haploid daughter nuclei with the same genotype as the parental nucleus. Meiosis occurs in reproductive cells. It begins with the formation of a diploid nucleus, and the division process gives rise to four haploid daughter nuclei. This is the sexual cycle, the outcome of which is the recombination and segregation of the genes that were brought together when the diploid was first formed. The progeny have new genotypes, but to achieve the expression of these the nuclei must be "packaged" into progeny spores that can be distributed into the environment. We will deal with the generalizations that simplify interpretation of meiotic segregations later, but we will start by describing the cell biology of sporulation in higher fungi.

In filamentous ascomycetes, hyphal fusion or similar mating between male and female structures results in nuclei moving from the male into the female to form an ascogonium in which male and female nuclei may pair but do not fuse (this is a dikaryon phase of limited extent). Ascogenous hyphae grow from the ascogonium. Most cells in these hyphae are dikaryotic, containing one maternal and one paternal nucleus, the pairs of nuclei undergoing conjugate divisions as the hypha extends. In typical development, the ascogenous hypha bends over to form a crozier. The two nuclei in the hooked cell undergo conjugate mitosis and then two septa are formed, creating three cells (Fig. 2.8). The cell at the bend of the crozier is binucleate, but the other two cells are uninucleate. The binucleate cell becomes the ascus mother cell, in which karyogamy takes place.

In the young ascus meiosis results in four haploid daughter nuclei, each of which divides by mitosis to form the eight ascospore nuclei (Fig. 2.8). Formation of ascospores results from the infolding of membranes around the daughter nuclei so that eight ascospores are delimited. A spore wall then forms around each ascospore. The spindle pole body (SPB; equivalent to the centrosome of other organisms) undergoes a duplication cycle in meiosis and mitosis. Using mutant analysis, it has been found that structural modification of the SPB in meiosis II is required to start endospore wall formation in *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*. Ascus cytoplasm left outside the spores, called the *epiplasm*, may provide nutrients to the maturing spores, contribute to the outer layers of the spore

wall, or contribute to the osmotic potential of the ascus to aid subsequent spore discharge.

Eight uninucleate and haploid ascospores are typically formed, although there are variations on this theme. In some species a further mitotic division forms binucleate spores; in others, the immature spores become multinucleate prior to being divided up by septa. Some ascomycetes form only four ascospores as a result of the spore membranes enclosing a pair of nuclei rather than just one nucleus, whereas others may produce fewer spores in each ascus as a result of nuclear disintegration or spore abortion.

In basidiomycetes, karyogamy and meiosis take place in the basidium, and basidiospores (usually four) are produced externally on outgrowths of the basidial wall, which are called *sterigmata* (Fig. 2.8). The walls of the sterigma and of the early spore initial are continuous and homologous. Starting with the spherical growth of the spore initial and nuclear migration into the maturing spore, a basidiospore grows further to attain the species-specific form and dimension, and the spore protection is increased by further wall layers formed with or without pigmentation, ornamentation, and a pore through which germination occurs.

All of these spore formation processes include steps that require precise nuclear positioning and active nuclear movement. In mushrooms and their relatives, the SPB always leads the daughter nucleus migrating into the maturing basidiospore. There are at least two nuclear movement systems in *Coprinus cinereus*. Microtubules are involved in the pairing of the two conjugate nuclei in the dikaryon. Migration of nuclei through clamp connections during formation of dikaryotic hyphae is dependent on tubulin, but migration of nuclei into developing spores is not.

4.5 Meiosis

Although there are some inevitable modifications, karyogamy and meiosis in heterothallic fungi go through stages fairly typical for eukaryotic haploids. In particular, the major round of DNA replication is premeiotic, occurring before karyogamy. Indeed, it was research with the ascomycete *Neotiella* that first demonstrated this aspect of meiosis. Meiosis takes place in the meiocytes (basidia or asci), and chromosome behavior in meiosis follows the Mendelian laws of segregation, independent assortment, linkage, and crossing over.

Particular landmarks in the meiocyte development pathway are: commitment to recombination, commitment to the first and second meiotic divisions (= haploidization), commitment to spore formation, and commitment to spore maturation. Chromosome pairing and synapsis are distinct processes in terms of both mechanism and timing. Premeiotic DNA replication and formation of the synaptonemal complex are independent events.

Synapsis does not require DNA homology. At prophase I, homologous chromosomes align before the tripartite synaptonemal complex appears. Double-strand breaks (providing sites for meiotic recombination) usually occur at this time. The fission yeast, *Schizosaccharomyces pombe*, has normal levels of meiotic recombination, but it does not make any synaptonemal complex, although structures resembling the axial core of the synaptonemal complex do occur. Mutants of *Saccharomyces cerevisiae* defective in synaptonemal complex structure show defects in chromosome condensation, but still undergo some meiotically induced homologous pairing, revealing the independence of homology pairing and meiotic recombination. In meiotic recombination, heteroduplex DNA, which refers to the hybrid-DNA formed following strand exchange containing one or more mismatched base pairs (see Chapter 6), is an essential intermediate. The frequency of homologous recombination is 100–1000 times higher during meiosis than during mitosis.

The synaptonemal complex, if formed, converts several sites of alignment and exchange into a functionally intact bivalent, consisting of kinetochores, that are sites for attachment to fibers of the division spindle, and the four DNA strands (the chromatids) connected by chiasmata. *Chiasmata* are the cytologically visible connections between homologous chromosomes. They have long been regarded as being equivalent to genetic crossovers formed in the recombination process. The number of chiasmata observed in meiosis of some organisms is much less than that of crossover events deduced from progeny analysis, however, so recombination events and chiasmata may not be numerically equivalent. Rather, chiasmata may be only those crossovers that are required to balance mechanical forces exerted by the division spindle on the kinetochores to ensure that homologues move away from each other in the first meiotic division.

Mutants defective in DNA repair have been demonstrated to have defects in meiotic chromosome condensation, synapsis, and recombination in meiosis, perhaps culminating in defects such as formation of inviable spores or abortion in sporulation. Using mutant analysis, it has been found that recombination events are not sufficient in themselves to ensure normal segregation of the chromosomes (called *disjunction*). On the other hand, experiments with artificially constructed chromosomes in *Saccharomyces cerevisiae* have shown that neither chromosome size nor homologous DNA sequence have much effect on chromosome disjunction. The crucial factor is that when a crossover occurs the chromosomes nearly always disjoin at meiosis I. Much more work clearly needs to be done to establish the precise relationships between, on the one hand, synapsis, sequence homology, and synaptonemal complex formation, and between molecular recombination events, genetic crossovers, and chiasmata on the other.

Unlike most plants and animals, fungi carry out meiosis with the nuclear membrane remaining intact in prophase I. Meiosis I is a reductional division and meiosis II is an equational division. The meiotic II division is mito-

sislike, sharing the same machinery with mitosis. At least in *Saccharomyces cerevisiae*, when proper environmental triggers are followed, the a1/α2 heterodimer formed from the mating-type gene products activates *IME*1 (Inducer of Meiosis) to synthesize a transcriptional factor that, in turn, switches on various meiotic genes. Yeast is the only fungus for which the physiological, biochemical, and molecular controls of meiosis are sufficiently well known for some understanding to emerge; it may or may not be representative.

The leptotene stage of prophase I is the stage during which chromosomes condense and are normally visible using conventional light microscopy, but this may be very brief or even absent in fungi. In addition, in contrast to plant and animal systems, fungal karyotype analysis by microscopy is usually done at the pachytene stage when chromosomes are paired and appear as long, thick threads, rather than at metaphase I, because fungal chromosomes are usually so small at the latter stage. A method of spreading and staining chromosomes with silver nitrate has been described that greatly improves conventional examination of the synaptonemal complex and chromosomal rearrangements by both light and electron microscopy. Fluorescence *in situ* hybridization (FISH technique) or "chromosome painting," has been applied to chromosome spreads of *Saccharomyces cerevisiae*. With a probe derived from the homologous genomic library, the technique allows the study of specific individual chromosomes during meiosis.

The karyotypes of most fungi can be resolved electrophoretically by pulsed field gel electrophoresis (PFGE) and southern hybridization. Southern hybridization with homologous probes can establish the ploidy levels of different isolates and reveal gene amplification during differentiation. This technique can also reveal the loss of supernumerary chromosomes, which are usually less than 1 million base pairs in size and are dispensable, and the generation of novel-sized chromosomes (chromosome length polymorphisms or CLPs, see Section 7.12). Chromosome length polymorphisms are widespread in both sexual and asexual species, revealing general genome plasticity. Tandem repeats (e.g., repeats of rRNA genes) frequently vary in length and dispensable chromosomes, and dispensable chromosome regions occur in fungal karyotypes. Many karyotype changes are genetically neutral; others may be advantageous in allowing adaptation to new environments.

4.6 Analyzing Gene Segregations from Random Spores

The outcome of meiosis is the formation of four haploid daughter nuclei. Chromosomes segregate at random during meiosis, so if the chromosomes carry detectable genes, then the frequencies of the genotypes among progeny spores that are analyzed at random can be easily predicted. For example, suppose you make an experimental cross involving two genes

located on different chromosomes by first making a heterokaryon between a strain carrying mutant *a*, and a compatible strain carrying mutant *b*. If you then give the heterokaryon the appropriate conditions for it to make a sexual fruiting body, a diploid will be formed. Remember that mutant strain *a* will carry the wild-type allele of mutant *b*, and mutant strain *b* will carry the wild-type allele of mutant *a*. Thus, the genotype of the diploid will be:

$$\frac{a}{+}\frac{+}{b}$$

When meiosis produces haploid progeny nuclei from this diploid, the daughter chromosomes carrying the two alleles of gene *a* will segregate independently of the daughter chromosomes carrying the two alleles of gene *b*, and four genotypes will arise: [*a*+], [*ab*], [+*b*] and [++]. Because the chromosomes *are* segregating independently, these four genotypes are formed in equal frequency, so a sample of the progeny population will contain 25% [*a*+], 25% [*ab*], 25% [+*b*] and 25% [++], which is often referred to as a 1:1:1:1 ratio.

That is the prediction, but, of course, the real world is not always as tidy as that. For one thing, you don't always know at the outset whether the genes are on the same chromosome or not. In such a case, you have to consider different possible outcomes and have some means of balancing your judgment between them. In the real world, you know that a sample may not always be fully representative of the sampled population, so you need a way of assessing how reliable a sample might be. Along with that, of course, because one fungal cross might generate several billion progeny spores, you cannot analyze all of them, so there must be some way of deciding how large a sample should be analyzed.

As an example, we can go back to the *Coprinus ade*-2 auxotroph. When that was being studied, progeny from a cross against wild type were scored for both adenine auxotrophy and mating type. The original diploid in this case had the genotype:

$$\frac{ade-2}{+}\frac{A_6}{A_5}$$

A total of 317 progeny were characterized, and they fell into the following four genotypes:

ade-2, A_6	74
+, A_5	72
ade-2, A_5	82
+, A_6	89

Those progeny numbers look fairly close to a 1:1:1:1 ratio, but the largest number is 89 and the smallest is 72, a full 17 different. Is that allowable?

Can we accept data sets that differ by so much as being "equal"? We need a statistical test that will provide guidance about the level of significance of the sorts of numerical deviations that occur in the real world. Most geneticists reach for the χ^2 significance test when such a need arises.

To explain this test, let's start with another, simpler, predicable situation: the tossing of coins. If a very large number (n) of coins were tossed together, equal numbers of heads and tails would be expected, but from common experience you would not expect to get $n/2$ heads in every experiment. Some variation would be expected from trial to trial. Indeed it is possible, though not very likely, when tossing several coins that the result would be all (n) heads. If four coins were tossed, you can use the relation $(p + q)^4$ to work out the possible outcomes of the toss (where p is the probability of a coin coming down heads, and q is the probability of the coin coming down tails). The general relationship $(p + q)^n$ can be used to work out the probabilities if n coins are tossed. These probabilities can be plotted graphically, and as n becomes very large, a smooth curve is obtained that approaches the continuous distribution called the *binomial distribution*. This type of curve applies to many biological variables.

From this distribution, another probability distribution, called *chi-squared* (χ^2) can be derived. By calculating the value of χ^2 for a given set of observations, one can discover how probable it is that the deviation from expectations is due to chance alone. For example, if the ratio of heads to tails from tossing a coin many times was far from 1:1, a χ^2 test would indicate that the likelihood of this happening by chance alone is remote for a normal coin. If there were only a remote probability that deviation from expectation is due to chance, one would have to suspect that the coin is biased in some way and more likely to fall with one side uppermost rather than the other. In other words, the χ^2 test helps in deciding whether the deviation between observation and expectation is not significant (deviations due to chance) or *significant* (in which case it must be due to a cause other than chance).

χ^2 is calculated from the formula: $\chi^2 = \Sigma[(o - e)^2/e]$. In this equation, "o" is the observed occurrence of the event; "e" is the expected occurrence of the event; and Σ means *the sum of* all such terms.

4.7 Use of χ^2 Tables

Calculation of χ^2 is relatively simple, but it's no use to you without the χ^2 tables. Such tables are to be found in almost all general books on statistics, and a short table follows. When you calculate a χ^2 value you are actually calculating the area under the distribution curve (hence, chi-*square*) to which the data corresponds. The χ^2 table tells you the probability value corresponding to that area. The probability you read off from the table is the

probability that chance alone will give rise to deviations as great as those between your observed and expected values.

High probabilities in the χ^2 test indicate very good agreement between observation and expectation (because they are saying that there is a high probability that deviations are due to chance; that is, to random fluctuations). On the other hand, low probabilities imply that it is unlikely that chance variation would produce such deviations. This of course means that if the deviations are not due to chance, then they must be due to something else that was not considered (or not known about) when the expectations were worked out.

You must recognize that your expectations constitute "a hypothesis" and the χ^2 test is your means of testing whether the observed data agree with that hypothesis. The χ^2 test unfortunately cannot explain your results on your behalf; you have to come up with an idea, convert it into numerical expectations, and then test those against experimental observations.

There clearly must be a borderline between high probabilities and low probabilities. The conventional border is set at the 5% probability level. It is usual to dismiss a hypothesis as unlikely to be true if the probability (that the deviations are due to chance) comes out less than 5%. In such a case the deviation is said to be significant. If the probability turns out to be more than 5% then the deviation is said to be "not significant" and the hypothesis on which the expectations were calculated is accepted. Five percent corresponds to the likelihood that such deviations will be observed once in every 20 trials; this is quite satisfactory as a rule, but for some purposes this may not be sufficiently stringent. You might set the borderline at 50%, which means that you'll accept something when chance would produce the deviations in about one trial in two; or even 95%, if the results were a matter of life and death. The χ^2 test is a tool for testing the validity of data; how you use it is up to you.

We show a short χ^2 table in Table 4.1. Note that such tables often, as here, give probability as a decimal fraction of 1; that is, 10% is given as 0.1, 5% as 0.05, and so on. Before you can use such a table, it is necessary to know the number of degrees of freedom, which the value in the extreme left hand column. This relates to the number of categories to which your data can be assigned. You will appreciate that there is a limit to the amount of really meaningful scientific research that can be done by tossing a coin around; real experiments very often have more than two categories. The number of categories, however, influences the shape of the distribution curve, so you need to know to which part of the χ^2 table to refer. It is quite simple, really, because it turns out that the number of degrees of freedom is normally one less than the number of separate classes or categories in the results.

It is easy enough to state; less easy to explain. In essence the number of degrees of freedom describes how many of the different options can be freely undertaken (i.e., entered into by chance). As soon as all but one of the options have been accounted for, there is evidently no further choice;

Table 4.1. A short table of χ^2 values.

	Probability (usually symbolized *P*)								
Degrees of freedom	0.99	0.90	0.80	0.50	0.20	0.10	0.05	0.02	0.01
1	0.000	0.016	0.064	0.455	1.64	2.71	3.84	5.41	6.64
2	0.020	0.211	0.446	1.386	3.22	4.61	5.99	7.82	9.21
3	0.115	0.584	1.005	2.366	4.64	6.25	7.82	9.84	11.35
4	0.297	1.064	1.649	3.357	5.99	7.78	9.49	11.67	13.28

thus, there are *n*-1 degrees of freedom, where *n* is the number of options, classes, or categories. Consider this example. If an individual datum can be either "big," "small," or "medium" (three categories), then there are only two degrees of freedom because an observation that is neither "big" nor "small" must be "medium." As a specific example, suppose you are told that a progeny of fungal spores was classified into four types; thus: [*ade*-2, A_6], [+, A_5], [*ade*-2, A_5], and [+, A_6]. If these are the only four phenotypes, having counted those that are of the [*ade*-2, A_6], [+, A_5], and [ade-2, A_5] types, then all the rest must be [+, A_6]. These data, therefore, have three degrees of freedom.

To use the χ^2 table in Table 4.1, suppose the calculated value of χ^2 for a set of data was found to be 2.5 with 3 degrees of freedom. Table 4.1 shows that this corresponds with a probability between 0.5 and 0.2. This is the range covering the probability that the observed data differ from the expected distribution by chance alone. In other words, the data certainly deviate from whatever ratio was predicted, but only to an extent that one could expect to occur by chance in as many as 50% of repeated experiments. Such a deviation is not exceptional and therefore gives grounds for accepting the hypothesis on which the original prediction was founded.

4.8 Testing for Homogeneity

The preceding approach uses the χ^2 distribution to test for what is known as *goodness-of-fit*, which is where a set of experimental data is being tested for compliance to some sort of expected or predicted distribution. The χ^2 test can also be used to establish whether two (or more) sets of data are similar to one another. There are many situations in genetics where you might be interested to know whether data from different experiments share the same distribution. This is a test of the homogeneity of the different sets of data, and it is possible to use a slightly more complex version of the χ^2 test to do this. For an example, let's go back to our coins. Suppose you tossed a coin 32 times and obtained 22 heads and 10 tails. This is significantly different from a 1:1 ratio (χ^2 value = 4.5 with one degree of freedom) and makes you suspicious about the trueness of the coin, so you try again. In the second experiment, of 20 tosses, there were 14 heads and 6 tails. This

deficiency of tails *looks* similar to the first test, but is not significantly different from a 1:1 ratio (χ^2 value = 3.2 with one degree of freedom). So, the first question to be asked is whether the two tests are homogeneous. For this, we carry out a χ^2 contingency test as follows.

First, draw up a table, known as a contingency table, of the observed results:

	Heads	Tails	Totals
Test 1	22	10	32
Test 2	14	6	20
Totals	36	16	52

From this, expected results are calculated based on the hypothesis that there is no difference between the tests. If there is no difference between them, then the separate test results will be distributed in the same way as the totals. As a result values in the totals cells can be used to calculate expected values. For example, the overall number of heads in the total is 36/52, but there were 32 tosses in Test 1; hence, the expected number of heads for Test 1 = (36/52) × 32 = 22.15. Repeating this process for each of the other three cells gives:

$$\text{expected heads Test 2} = (36/52) \times 20 = 13.85,$$

$$\text{expected tails Test 1} = (16/52) \times 32 = 9.85,$$

$$\text{expected tails Test 2} = (16/52) \times 20 = 6.15$$

We now have a full set of observed and expected values and can calculate χ^2 using the equation quoted earlier, $\chi^2 = \Sigma[(o - e)^2/e]$:

$$\begin{aligned}&\left[(22-22.15)^2/22.15\right]+\left[(14-13.85)^2/13.85\right]\\&\quad+\left[(10-9.85)^2/9.85\right]+\left[(6-6.15)^2/6.15\right]\\&=0.0010+0.0016+0.0023+0.0036\\&=0.0085\end{aligned}$$

A 2 × 2 contingency table has only one degree of freedom (if we set out a table with nothing but the totals shown, only *one* of the blank spaces needs to be filled to determine what should be in all the other three). Hence, there is one degree of freedom. For a value of $\chi^2 = 0.0085$, therefore, P is greater than 0.90. That is, there is a greater than 90% probability of getting this sort of deviation by chance. We would conclude that the data show no significant difference between the two tests. The value of this is that if there is no significant difference between the tests, we can add the two sets of data

together and use the combined data as a much larger sample size to show that its ratio of 36 heads to 16 tails is very significantly different from 1:1 (χ^2 value = 7.7 with one degree of freedom, P much less than 0.01).

There are two very important points about χ^2 tests. First, they should not be used if any of the expected values fall below 5 or if one or more class is absent from the observed data. Second, all of the calculations must be carried out on the actual numbers observed or expected; percentages or other proportions *must not be used* because the significance of a deviation depends on the number of observations made. Comparison of the following two examples emphasizes this point. A coin is tossed 20 times, giving 7 heads and 13 tails. Does this result differ significantly from the expected 1:1 ratio?

	Heads	Tails
o	7	13
e	10	10
o – e	–3	3
$(o - e)^2$	9	9
$(o - e)^2/e$	0.9	0.9

$\chi^2 = 0.9 + 0.9 = 1.8$; from the χ^2 table (Table 4.1), the probability of getting a χ^2 value like this is around 10%. This means that chance variation alone will produce deviations of this extent from the expected results in at least one trial in every 10. This is quite acceptable, and we are willing to accept the coin as "true" and say that the observed ratio of 7 heads to 13 tails is not significantly different from the expected 1:1 ratio.

Suppose, however, that the coin was tossed 200 times, giving 70 heads and 130 tails (i.e., keeping to the same *ratio* as before). The calculation now comes out rather differently:

	Heads	Tails
o	70	130
e	100	100
o – e	–30	30
$(o - e)^2$	900	900
$(o - e)^2/e$	9	9

This time, $\Sigma\chi^2 = 18$, which corresponds to a probability very much less than 1%. In other words, chance variation alone would produce such deviations only once in every hundred or more trials. This is too low a probability to be acceptable, so we say the observed result is significantly different from expected. In this case, then, we have to re-examine the assumptions on

which the expectation was based; that is, that the coin is "true" and equally likely to come down heads or tails when tossed. This coin clearly comes down tails more often than heads; it could be weighted on one side, or irregularly shaped, or perhaps it's being tossed unfairly. We don't know *why* it's peculiar, but we have detected that it *is* peculiar.

Notice, though, that these two examples used the same observed ratio of heads to tails. The χ^2 test distinguished between them because it deals with numbers rather than ratios or percentages. Notice also the power of a large sample size. The minimum sample size required to distinguish with certainty between two possible ratios can be calculated using this equation:

$$n = \left[\frac{1 + (ab)^{\frac{1}{2}}}{a^{\frac{1}{2}} - b^{\frac{1}{2}}} \right] \chi^2_P$$

Where a = the first ratio,
b = the second ratio,
and χ^2_P = the χ^2 at the desired level of probability, P.

You can use this equation to calculate n for either end of a range of values you need to distinguish. For example, to distinguish between a 1:99 ratio and a 5:95 ratio with 95% probability needs a sample size of 169; to distinguish 30:70 from 40:60 with 95% probability requires a sample size of 244. Again, you have to deal with *numbers*, and the numbers must be realistic to your experimental situation.

We are now equipped with the necessary techniques to answer the question raised some pages ago about the segregation of the *Coprinus ade*-2 auxotroph and the *A* mating-type locus. Remember, the progeny numbers look fairly close to a 1:1:1:1 ratio, but now we can use the χ^2 test to determine just how close. Thus, the proposition we are testing is that the progeny show random segregation and that they consequently fall into a 1:1:1:1 ratio. The χ^2 calculation looks like this:

	ade-2, A_6	+, A_5	*ade*-2, A_5	+, A_6
o	74	72	82	89
e	79.25	79.25	79.25	79.25
o − e	5.25	7.25	2.75	9.75
$(o - e)^2$	27.56	52.56	7.56	95.06
$(o - e)^2/e$	0.35	0.66	0.09	1.20

From this calculation, $\Sigma\chi^2 = 2.3$; there are 3 degrees of freedom, so this corresponds to a probability close to 0.5 that this data set is *not* significantly different from a 1:1:1:1 ratio. There is strong evidence, therefore, that these genes are unlinked (i.e., probably located on different chromosomes).

You will appreciate that if you are using mating-type tester strains to determine the *A* mating-type specificity of a set of progeny of *Coprinus*, the same tests can be use to determine the *B* mating-type specificity. Thus, the analysis that yielded the *ade*-2 and *A* mating-type progeny phenotypes we have just analyzed also provide the following data for segregation of *ade*-2 and the *B* mating-type locus, which we can also test for compliance with a 1:1:1:1 ratio:

	ade-2, B_6	+, B_5	*ade*-2, B_5	+, B_6
o	82	81	74	80
e	79.25	79.25	79.25	79.25
o – e	2.75	1.75	5.25	0.75
$(o - e)^2$	7.56	3.06	27.56	0.01
$(o - e)^2/e$	0.1	0.04	0.35	1.20

From this calculation, $\Sigma\chi^2 = 1.69$; again, there are 3 degrees of freedom, and this corresponds to a probability between 0.5 and 0.8 that this data set is *not* significantly different from a 1:1:1:1 ratio. Once more, there is strong evidence that these genes are unlinked, and probably located on different chromosomes. Thus, with a single cross we have established that *ade*-2, the *A* mating-type locus, and the *B* mating-type factor of *Coprinus cinereus* segregate independently of one another and are consequently probably located on different chromosomes.

Another, similar, example from *C. cinereus* is a cross that involves the methionine auxotroph *met*-5. The original cross was between a strain of genotype *met*-5, A_6, B_6, and a wild type of mating-type A_5, B_5. A sample of progeny was tested for auxotrophy and for both mating-type specificities. The *met*-5/*B*-mating type segregation was as follows:

	met-5, B_6	+, B_5	*met*-5, B_5	+, B_6
o	102	75	81	101
e	90	90	90	90
o – e	12	15	9	11
$(o - e)^2$	144	225	81	121
$(o - e)^2/e$	1.6	2.5	0.9	1.34

In this case, $\Sigma\chi^2 = 6.34$; again, there are 3 degrees of freedom, and this corresponds to a probability between 0.05 and 0.1 that these deviations are due to chance variation and that this data set is *not* significantly different from a 1:1:1:1 ratio. These two genes are probably located on different chromosomes.

Finally, the segregation of *met*-5 with the *A*-mating type character was:

	met-5, A_6	+, A_5	*met*-5, A_5	+, A_6
o	128	125	56	51
e	90	90	90	90
o – e	38	35	34	39
$(o - e)^2$	1444	1225	1156	1521
$(o - e)^2/e$	16.0	13.6	12.8	16.9

Now, $\Sigma\chi^2 = 59.3$, which, with 3 degrees of freedom, is far to the right of the values shown in the χ^2 table in Table 4.1, so it corresponds to a probability far less than 0.01 that this data set differs significantly from a 1:1:1:1 ratio simply due to chance. If the difference is not due to chance, it must be due to something else; indeed, this result provides evidence for these two genes being located on the same chromosome.

4.9 Detecting Linkage

To explain why we conclude that *met*-5 and mating-type factor *A* are on the same chromosome, let's go back to look at the exact genotype of the diploid used in the cross. The original cross was between a strain of genotype *met*-5, A_6, B_6 and a wild type of mating-type A_5, B_5. We can ignore mating-type factor *B* now because we have evidence for its independent segregation. The original diploid genotype that concerns us looked like this:

$$\frac{met-5}{+}\,\frac{A_6}{A_5}$$

Note that [*met*-5, A_6] were brought into the cross together from one parent, and [+, A_5] were brought in together from the other parent. These two are the *parental genotypes*. If you look at the observed progeny genotypes from the cross in the χ^2 calculation preceding table, you will see that these two genotypes predominated among the progeny, accounting for more than two thirds of them; out of 360 progeny, 128 were [*met*-5, A_6] and 125 were [+, A_5]. The other progeny were 51 of genotype [+, A_6] and 56 of genotype [*met*-5, A_5]; because these have their genes in reassorted arrangements compared with the parents they are called *recombinant genotypes*.

If chromosome segregation is the only thing that happens to the chromosomes carrying the [*met*-5, A_6] and [+, A_5] combinations of alleles as the diploid nucleus goes through meiosis, then only these parental genotypes can occur among the progeny. We have seen, however, that a proportion of progeny with recombinant genotypes do occur, so recombination events (also called *crossovers* or *chiasmata*) occur between the two genes in addition to chromosome segregation. As we will illustrate in the Chapter 5, it can be demonstrated that all four genotypes (two parental and two recombinant genotypes) can emerge from a single meiosis. This indicates that

recombination (or crossing over) occurs at the four-strand stage of meiosis and involves segments of DNA being exchanged between nonsister chromatids of homologous chromosomes.

A significant feature of recombination is that it occurs regularly. Its rate of occurrence is calculated as a percentage using the equation:

$$\frac{\text{Number of recombinant progeny}}{\text{Total progeny analysed}} \times 100\%$$

For the *met*-5 and *A* cross detailed before this equation turns out like this:

$$\frac{107}{360} \times 100\% = 29.7\%$$

So the recombination frequency between *met*-5 and *A* is 29.7%. Experience shows that, although there will be some statistical variation, whenever you make a cross in *Coprinus* involving these two genes, you will find a recombination frequency of about this level. This implies that the recombination frequency is a characteristic of the two genes involved in the cross. If this is true, then it gives rise to an assumption that permits the relative positions of genes on a chromosome to be deduced. That assumption is that crossing over is a random event (i.e., there is an equal chance of it happening at any position along a pair of nonsister chromatids at the four-strand stage of meiosis). If this is assumed then it follows that the recombination frequency you can measure in genetic crosses can be used as a measure of distance between the genes.

The argument runs in this way. If crossing over is randomized along the length of a chromosome, then it will occur more often between two genes that are far apart than it will between two genes that are close together. In other words, recombination frequency will be proportional to the distance between genes. Taking this argument further, if you make the right set of crosses you could calculate the recombination frequencies (also known as linkage distances) for different pairs of genes and construct a map of their relative positions on the chromosome.

Table 4.2 shows the results of two more crosses involving *Coprinus* that illustrate this. You can calculate the exact recombination frequencies for yourself, but you will find that the *ade*-8/*A* interval has a recombination frequency of just more than 1%, and the *ade*-8/*met*-5 interval is around 30%. Because we know that *A*/*met*-5 is 29.7%, we should be able to establish the gene order. It is obvious that *ade*-8 and *A* are very close together, and they both show about 30% recombination with *met*-5, but to distinguish the order *ade*-8–*A*–*met*-5 from the alternative order *A*–*ade*-8–*met*-5 with two-point crosses is difficult. It is difficult because we would have to distinguish between two very similar recombination frequencies; let's say we'd have to distinguish 29% and 30% from 30% and 31%. The standard error (SE) of recombination frequency depends on sample size, and the fact is, except

Table 4.2. Results of two additional crosses involving genes linked to the *A* mating-type factor of *Coprinus cinereus*.

1. Cross between *ade*-8 and *A*	*ade*-8, A_6 105	+, A_5 111	*ade*-8, A_5 2	+, A_6 1
2. Cross between *ade*-8 and *met*-5	*met*-5, *ade*-8, 57	+, + 66	*met*-5, + 132	+, *ade*-8 136

with really very large sample sizes, the SE on recombination values in the region of 30% is likely to be too great for reliably distinguishing these two gene orders. A more efficient way of establishing the order of three genes is to have all three in one cross, the so-called three-point cross. The progeny types obtained from a three-point cross reveal the order of the genes as well as their distances.

This sort of linkage analysis will be considered in detail in the Chapter 5. We will close this chapter by saying that enough comparisons have now been made between linkage maps of this sort and the actual physical DNA sequence to show that the two match tolerably well. It turns out that crossing over is not entirely a random event. There are regions of DNA where crossing over is suppressed, and there are other regions that are more likely than normal to be involved in recombination (these are called *recombination hotspots*). The result is that a recombination "distance" on a linkage map may not be a direct measure of the physical distance between two genetic sites. Nevertheless, linkage maps have generally proved to be sufficiently accurate to provide satisfactory planning guidelines for breeding programs, and even genome sequencing programs.

Publications and Websites Worth a Visit

Atherly, A.G., Gorton, J.R. & McDonld, J.F. (1999). *The Science of Genetics*. Saunders College Publishing, Harcourt Brace & Co.: New York.

Davis, R.H. (2000). *Neurospora: Contributions of a Model Organism*. Oxford University Press: Oxford, U.K.

Esser, K. (1996). Meiotic recombination. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 43–71. Marcel Dekker, Inc.: New York.

Fungal Genetics Stock Center website at http://www.fgsc.net/methods.

Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C. & Gelbart, W.M. (2000). *An Introduction to Genetic Analysis*, Seventh edition. W.H. Freeman and Company: New York. [and visit http://www.whfreeman.com/genetics]

Hartwell, L., Hood, L., Goldberg, M.L., Reynolds, A.E., Silver, L.M. & Veres, R.C. (2000). *Genetics: From Genes to Genomes.* McGraw-Hill Inc: New York. [and visit http://www.mhhe.com/hartwell]

Lu, B.C.K. (1996). Chromosomes, mitosis, and meiosis. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 119–176. Marcel Dekker, Inc.: New York.

Martinelli, S.D. & Kinghorn, J.R. (1994). *Aspergillus: 50 Years On*. Elsevier: Amsterdam.

Novick, P., Ferro, S. & Schekman, R. (1981). Order of events in the yeast secretory pathway. *Cell* **25**, 461–469.

Perkins, D.D., Radford, A. & Sachs, M.S. (2000). *The Neurospora Compendium: Chromosomal Loci.* Academic Press: Orlando, FL.

Statpages.net is a website that collects together interactive pages that perform statistical calculations. There are also links to books, tutorials, software and other resources related to statistical analyses.

Strathern, J.N., Jones, E.W. & Broach, J.R. (1981). *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. Cold Spring Harbor Laboratory Press: New York.

Winter, P.C., Hickey, G.I. & Fletcher, H.L. (1998). *Instant Notes in Genetics*. Bios Scientific Publishers: Oxford, U.K.

Historical Publications Worth Knowing About

Beadle, G.W. & Tatum, E. (1945). *Neurospora.* II. Methods of producing and detecting mutations concerned with nutritional requirements. *American Journal of Botany* **32**, 678–686.

Clutterbuck, A.J. (1974). *Aspergillus*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 447–510. Plenum Press: New York.

Esser, K. (1974). *Podospora anserina*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 531–551. Plenum Press: New York.

Gans, M. & Prud'homme, N. (1976). Chromosome maps of *Coprinus radiatus*. In *Handbook of Biochemistry*, Third edition, *Section B, Nucleic Acids*, vol. 2, (G.D. Fasman, ed.), pp. 762–764. CRC Press: Cleveland.

Moore, D. (1967). Four new linkage groups in *Coprinus lagopus*. *Genetical Research* **9**, 331–342.

Moore, D. (1967). Purine-requiring auxotrophs of *Coprinus lagopus* (*sensu* Buller). *Journal of General Microbiology* **47**, 163–170.

Mortimer, R.K. & Hawthorn, D.C. (1966). Genetic mapping in *Saccharomyces*. *Genetics* **53**, 165–173.

Radford, A. (1976). Chromosome maps of *Neurospora crassa*. In *Handbook of Biochemistry*, Third edition, *Section B, Nucleic Acids*, vol. 2, (G.D. Fasman, ed.), pp. 739–761. CRC Press: Cleveland.

Sherman, F. & Lawrence, C.W. (1974). *Saccharomyces*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 359–393. Plenum Press: New York.

Westergaard, M. & von Wettstein, D. (1970). Studies on the mechanism of crossing over. IV. The molecular organization of the synaptinemal complex in *Neotiella* (Cooke) Saccardo (Ascomycetes). *Comptes Rendus des Travaux du Laboratoire Carlsberg* **37**, 239–268.

CHAPTER 5

Recombination Analysis

Revision Concepts

- Random assortment is indicated by 50% recombination frequency: two genes are either on different chromosomes or on the same chromosome, but very far apart.
- Recombination frequency in an interval between two loci represents the sum total of all crossovers in that interval.
- Progeny from a three-point cross always fall into eight categories: parental genotypes are always in the majority, whereas double recombinants form the smallest size class.
- Progeny frequencies can be used to determine the order of the genes as well as to calculate recombination frequencies.
- Crossing over in one region is not independent of crossing over in the region adjacent to it, which is a phenomenon called *chiasma interference*, and is measured by the coefficient of coincidence.
- The percentage of meioses in which a crossover occurs equals twice the recombination frequency.
- Only fungi allow the ability to map the position of the centromere because meiotic products are kept together in the ascus (usually an octad of eight ascospores) or on the basidium (usually a tetrad of four basidiospores).
- Ordered octads occur in some ascomycetes with asci so narrow that division spindles are always in the long axis.
- Unordered octads (occasionally tetrads) occur in other ascomycetes and unordered tetrads occur in basidiomycetes.
- The distance between a gene and its centromere can be calculated directly by recording the number of second division ascospore segregations in ordered octads.
- Ascus segregations can be used to study linkage between genes: asci segregate into classes of parental ditype (PD), nonparental ditype (NPD), and tetratype (T) octads or tetrads.
- With unlinked genes, PD and NPD arise in equal frequency and T arise when one gene recombines with its centromere; with linked genes,

PD are frequent, NPD only arise by four-strand double crossover and are rare, and T arise from single crossovers between the two genes.

- In tetrad analysis a PD:NPD ratio of 1 is absolutely diagnostic of no linkage.
- Difficulties in tetrad analysis include isolating spores due to their dense distribution, assessing ripe and unripe tetrads, poor spore germination, and high losses due to contamination.
- Unordered tetrad analysis requires a centromere marker or at least two other genes, one of which is independent of the gene whose centromere distance is to be determined.
- The segregation rules are the same for unordered and ordered tetrads.
- Maps constructed using genetic segregations represent the behavior of genes and chromosomes as they progress through the meiotic division cycle.
- Some organisms use the mechanics of meiosis to control their sexuality, becoming secondarily homothallic by combining more than one nucleus per spore.
- Mitotic segregations (or the parasexual cycle) can be used to map chromosomes.
- Nutritional selection or spore color selection can isolate diploid spores, expressing a heterozygous phenotype from the many haploid spores that cannot.
- Haploidization is the process of chromosome loss during successive aberrant mitoses as a result of improper transport of chromosomes to the poles of the division spindle.
- Genes on the same chromosome show complete linkage during haploidization, whereas genes assort freely if not on the same chromosome.
- Mitotic homologous chromosomes do not form a synaptonemal complex, but occasional reciprocal recombination can occur.
- In mitotic diploid segregants, homozygosity for one gene will always be accompanied by homozygosity of any genes distal to it on the same chromosome, but not necessarily by homozygosity of genes proximal to it.
- Mitochondrial genomes specify some organelle proteins, with the rest being specified by nuclear genes.
- Vertical mitochondrial transmission can be female/male dependent, whereas horizontal transmission results from hyphal fusion and is characterized by cytoplasmic segregation.
- Mitochondrial inheritance is manifested in a wide range of phenotypes, from nonmigration to sectoring or mosaics to mixing, which results in recombination between mitochondrial genes.
- Mitochondria contain some plasmids that are neutral, but others that can be deleterious.
- Viruslike particles in fungi are similar to RNA viruses and are transmitted through hyphal fusion.

- Yeasts carry retroviruslike elements, or transposons, that integrate into chromatin, but are transmitted in nonmendelian fashion.
- Fungi have prion proteins that redirect folding of native polypeptides into abnormal configurations resulting in nonfunctional proteins.

5.1 Linkage Studies Make Maps

If the recombination frequency is a measure of distance between two genes, it should be possible to carry out a systematic series of crosses that will give you the recombination frequencies for pairwise combinations of a collection of different genes. From these, you should be able to construct a map of the relative positions of the genes on the chromosome. This proves to be perfectly feasible, but there are a few points we need to make before illustrating the operation.

The first point is that there is a limit to the amount of recombination that can be recognized in a single cross. That limit is set by the outcome of a cross between two genes that are known to be located on different chromosomes. We have already explained in Chapter 4 that a cross between mutant strain *a* and mutant strain *b* will produce a diploid with the genotype:

$$\frac{a}{+}\,\frac{+}{b}$$

When such a diploid goes through meiosis, it will give rise to progeny of four genotypes in equal frequency; thus, a sample of the progeny population will contain the following genotypes:

$$25\%\,[a+]\,(\text{parentals})$$

$$25\%\,[+b]\,(\text{parentals})$$

$$25\%\,[ab]\,(\text{recombinants})$$

$$25\%\,[++]\,(\text{recombinants})$$

The recombinant progeny obviously amount to 50% of the total, yet because we set this feature at the outset we know that these genes are on different chromosomes. A consequence of this is that if we cross two unknown genes together and get a close to 50% recombinant genotypes we can only describe them as showing "random assortment" or "random segregation." The two genes might be on different chromosomes and genuinely "unlinked," but on the other hand they might be on the same chromosome (that is "linked") but so far apart that crossover occurs between them in all meioses. Yes, that's right; one crossover somewhere between two linked genes occurring in each meiosis will give you 50% recombination.

Table 5.1. Some imaginary recombination values.

Genes in the cross	Recombination %
$x \times z$	18
$x \times y$	15
$y \times z$	5
$w \times z$	43
$w \times y$	45

Thus, 50% recombination is the upper limit, but statistical variation comes into this. We can use the equation in Section 4.8 to calculate the minimum number of progeny required to distinguish between, say, 45% recombinant progeny and 50% recombinant progeny with different degrees of certainty. For 95% probability you will need to analyze a total of 1563 progeny. If you want to be 99% certain, you will have to analyze at least 2700 progeny. This means that in normal practice, where only a couple of hundred progeny might be routinely analyzed, recombination values in the 40% region cannot be distinguished from random segregation with any great certainty.

Despite this limitation, it is feasible to intercross a collection of mutants and arrange them into linkage groups (i.e., subsets of mutant strains which, in mutant × mutant crosses, give a recombination percentage that is in statistical terms significantly less than 50%. The recombination percentages can then be sorted into linkage group maps. For example, consider the small collection of imaginary recombination values shown in Table 5.1. The first three results contribute to a reasonable map, with the genes in the order *x*–*y*–*z*, as shown in Fig. 5.1. This little map is quite consistent with the data we have for these genes and, although it is not perfect, it is acceptably additive (15 + 5 approximates satisfactorily to 18). Bear in mind that in practice recombination frequency measurements such as these might have standard errors of about 10% of the value stated in the table, in which case this additivity equation would be more realistically stated as [(15 ± 1.5) + (5 ± 0.5) = 18 ± 1.8]. We are therefore reasonably satisfied with the map for these three genes.

That is all well and good, but what about gene *w*? With the data we have in the table, there are three possibilities: (1) unless a very large sample of progeny was analyzed, there's a possibility that 45% and 43% recombina-

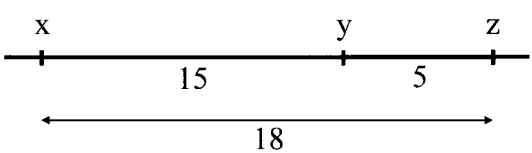


Fig. 5.1. A simple, and imaginary, three-point chromosome map.

tion are not significantly different from random segregation, so gene *w* might not even be linked to the other three; (2) taking the 45% and 43% recombination frequencies at face value, gene *w* could be far to the right of gene *z*; but (3) equally, gene *w* could be far to the left of gene *x*. We need more data to position gene *w*, but whatever its location it will be difficult to position reliably. We might consider making a $w \times x$ cross in the expectation that the result would indicate random segregation if gene *w* is to the right of *z*, but a recombination value of around 30% if gene *w* is to the left of gene *x*. Remembering the variability expected in normal experiments, however, we would have to analyze a large number of progeny, probably from several hundred to a couple of thousand, if we want to obtain a believable result, but there is a better way.

Much more information is obtained from a cross if three rather than two genetic markers are followed. In three-point crosses, recombination frequencies are obtained in the usual way, but they also allow the relative order of the genes on the chromosome to be determined, just by direct inspection of the data. To see why this is so we will start by using the basic knowledge we already have about recombination to predict the outcome of an imaginary three-point cross. We should then be able to recognize the underlying rules that will allow us to interpret the results of real-life crosses.

5.2 Multipoint Crosses

For our imaginary cross, we will use the imaginary genes *a*, *b*, and *c*. We are going to assume that they are all linked together with 10% recombination between *a* and *b*, and 10% recombination between *b* and *c*. The map for this region is shown in Fig. 5.2a. We are planning to cross a strain carrying mutations in all three genes (a triple mutant strain) with the wild type, so the heterokaryon will be [*abc*] + [+++] and the parental diploid nucleus, just before the start of meiosis, will have the appearance shown in Fig. 5.2b.

The first task is to predict the full range of genotypes that meiosis can produce and which are expected to appear among the progeny in consequence. The simplest thing that can happen during meiosis is that the homologous chromosomes replicate and the daughter chromatids disjoin without recombining. When that happens, progeny of parental genotypes are produced; namely, *abc* and +++.

In some meiotic divisions (about 20% of them, but we will deal with the frequencies later), however, a recombination event will occur between genes *a* and *b* at the four-strand stage (Fig. 5.2c), giving rise to the recombinant genotypes +*bc* and *a*++.

In other meiotic divisions (again, about 20%) a recombination event similarly will occur between genes *b* and *c* at the four-strand stage (Fig. 5.2d). This event will produce the recombinant progeny genotypes *ab*+ and ++*c*.

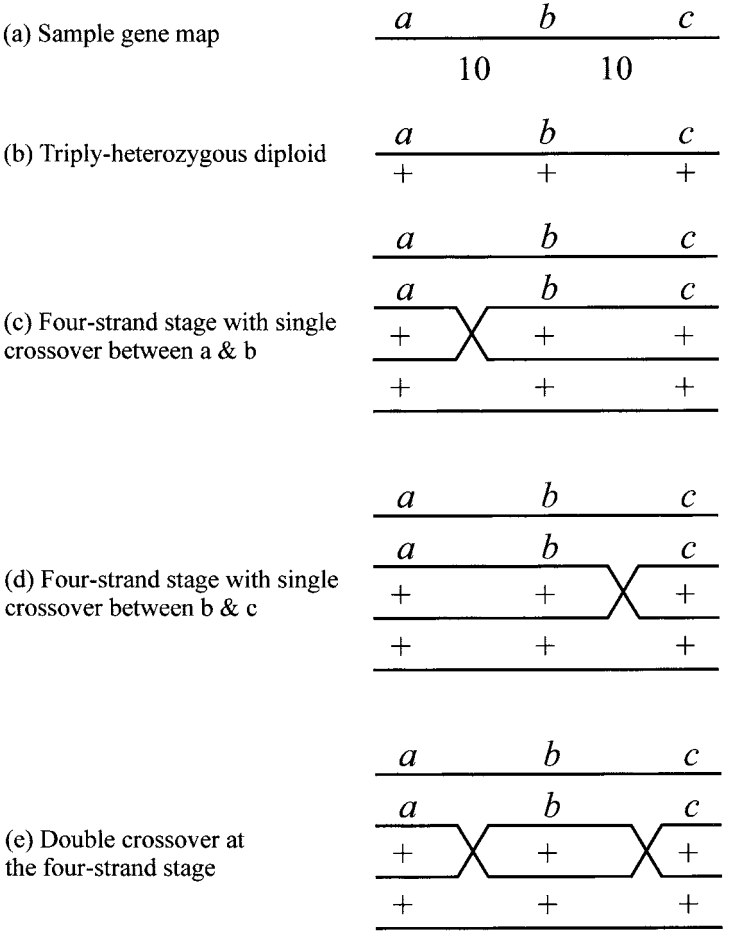


Fig. 5.2. Predicting the outcome of a three-point cross. (a) The disposition of the genes; (b) full genotype of the triply heterozygous parental diploid nucleus, just before the start of meiosis; (c) a recombination event between genes *a* and *b* at the four-strand stage; (d) a recombination event between genes *b* and *c* at the four-strand stage; (e) both of these recombination events occurring at the four-strand stage.

Remember that recombination mapping like this assumes that recombination events occur at random. Now, if recombination between *a* and *b* is occurring at random, and recombination between *b* and *c* is also occurring at random, then there must be some meiotic divisions in which *both* of these recombination events occur at the four-strand stage (Fig. 5.2e). This

Table 5.2. Summary, origins of genotypes, and predicted frequencies of progeny expected from the imaginary cross illustrated in Fig. 5.2.

Genotype	Nature and/or origin	Frequency (%)
abc	Parental, formed by chromatid disjunction only	40.5
+++	Parental, formed by chromatid disjunction only	40.5
a++	Single crossover in interval I (between *a* and *b*)	4.5
+bc	Single crossover in interval I (between *a* and *b*)	4.5
ab+	Single crossover in interval II (between *b* and *c*)	4.5
++c	Single crossover in interval II (between *b* and *c*)	4.5
a + c	Double crossover	0.5
+b+	Double crossover	0.5

situation will produce the recombinant progeny genotypes *a+c* and *+b+*. We have now predicted all the expected genotypes, and they have been brought together into Table 5.2.

We can use the information we have about recombination frequencies to calculate the expected frequency of each of these progeny types, starting with the double crossovers. We know that there is 10% recombination between *a* and *b*, and 10% recombination between *b* and *c*. Now, if recombination really does occur at random, then a recombination frequency is a probability statement. That is, 10% recombination is equivalent to saying that there is a 10% probability of a recombination event occurring between *a* and *b*, and, in this case of course, there is also a 10% probability of a recombination event occurring between *b* and *c*. Two truly random events occur independently of one another. That is, if event *x* is really random, then it must occur without reference to a separate event, *y*, and the probability of both events occurring together is the algebraic product of their separate probabilities. In our example, that works out to 10% of 10%, which is 1%.

Be careful when you are calculating with percentages to remember that they are proportions, not whole numbers. It is probably safer to convert them to decimal fractions of 1 for calculation, so that 10% of 10% becomes $0.1 \times 0.1 = 0.01$. Whichever route of calculation you use, the end product is that 1% of the progeny will arise from double crossovers. Because crossovers are reciprocal, the two genotypes that result from double crossovers will each form half the total: there will be 0.5% *a+c*, and 0.5% *+b+*.

For the single crossover genotypes it's not as simple as assuming that their frequency is the same as the recombination frequency. The double crossovers provide the complication. Each double crossover progeny, by definition, represents a crossover in each interval. A recombination frequency of 10% in an interval means that the sum total of all crossovers, single and double, amounts to 10%. It follows, then, that the expected frequency of single crossovers in interval I (between *a* and *b*) is 10% (the recombination frequency) minus 1% (the expected frequency of double

crossovers) = 9%. Again, the crossovers are reciprocal, so the two genotypes that result from this crossover will each form half the total: there will be 4.5% +*bc* and 4.5% *a*++. To ease the calculation in this example we made the recombination values the same in the two intervals; that's not likely to be the case very often, of course, but right now the expected frequency of single crossovers in interval II (between *b* and *c*) is 10%, subtract 1% (the expected frequency of double crossovers is used again because they do, after all, have two crossovers), resulting in 4.5% *ab*+ and 4.5% ++*c*.

To work out the expected frequency of progeny with the parental genotypes, total up the recombinants we've just calculated and subtract from 100%. The complete set of expected progeny frequencies are listed in Table 5.2. This completes the predictions for the progeny from this cross. Now let's extract the general rules for analyzing three-point crosses.

5.3 Rules of the Three-Point Crosses Game

Progeny from a three-point cross always fall into eight genotypes, which are in four pairwise combinations of reciprocal genotypes. When the three genes in the cross are linked, the progeny types appear in a very characteristic pattern of frequencies.

- One pair of genotypes is always in a large majority. These are the parental genotypes. They reveal the exact genotypes of the strains used to make the cross and, consequently, the exact genotype of the triply heterozygous diploid nucleus.
- One pair of genotypes is always in a small minority. These are the genotypes that arise from double recombinants and they are always the least frequent classes of progeny because they do require two recombination events.
- There are then two pairs of genotypes that arise from single recombinants. The numbers of progeny in these classes depends on the distances between the genes on the chromosome. A large interval between two genes will contain many crossovers and produce many singly recombinant progeny. A small interval will contain few crossovers and correspondingly few singly recombinant progeny. If the two intervals are about the same size, then these classes will also contain about the same number of progeny, as they do in our theoretical example, earlier.

The progeny frequencies can obviously be used to calculate the recombination frequencies and map the genes, but there is an important piece of information that can be obtained from inspection of the data before the calculation starts: the order of the genes. Three genes can be arranged in three different orders (e.g., *a*–*b*–*c*, *a*–*c*–*b*, or *b*–*a*–*c*). In most cases when you set out to perform the cross you will not know the order of the genes in advance; it's one of the pieces of information you need to get from the

results of the cross. You can infer the order by comparing the genotypes of the parentals with the genotypes of the double recombinants. This is because the parental genotypes show you the chromosomal arrangements that entered the meiosis, and the double recombinants reveal the chromosomal arrangements that result from the original after two crossovers, one on either side of the middle gene. It is specifically the central marker in the set of three that the two recombination events re-arrange. The trick, therefore, is to compare the parental genotypes with the double recombinant genotypes; only one gene will have changed its arrangement relative to the others, and that gene is the one that's in the middle.

Look back at our theoretical example; *abc* and +++ occur in the greatest frequency and are the parentals. The least-frequent genotypes are *a*+*c* and +*b*+; these are the double recombinant genotypes, and comparing them with the parentals (compare *a*+*c* & *abc* and +*b*+ & +++) shows that gene *b* is in the middle of the three. That was easy; we knew that from the start! Let's try a genuine sample three-point cross.

5.4 A Three-Point Cross in *Coprinus*

This experiment used three nutritional mutants, *met*-1, a methionine auxotroph, *ade*-2, an adenine-requiring mutant and *pdx*, an auxotroph that cannot synthesize pyridoxine (vitamin B_6; used as a coenzyme for many enzymes, especially transaminases and decarboxylases). The cross was set up between the double mutant strain *pdx*, *ade*-2, and the *met*-1 isolate. The basidiospore viability was a satisfactory 87%, and a total of 411 progeny were analyzed, yielding the progeny genotype frequencies shown in Table 5.3. The numbers observed in the progeny classes corresponding to the genotypes *pdx ade*-2 + and + + *met*-1 confirm the (given) parental genotypes. It is equally evident, in terms of the numbers observed, that + *ade*-2 + and *pdx* + *met*-1 are the double recombinants. Comparing these genotypes shows that *pdx* is the central gene of the trio. Compare parental + + *met*-1 with double recombinant *pdx* + *met*-1; or compare parental *pdx*

Table 5.3. Progeny from a three-point cross in *Coprinus cinereus*.

Progeny genotype	Number
pdx, *ade*-2, +	155
+, +, *met*-1	170
pdx, +, +	36
+, *ade*-2, *met*-1	30
pdx, *ade*-2, *met*-1	7
+, +, +	10
pdx, +, *met*-1	1
+, *ade*-2, +	2

ade-2 + with double recombinant + *ade*-2 +; in each case *pdx* has changed its relationship with the other two genes; it has been recombined. Because it is the central gene that recombines in double recombinants, it follows that the gene map must be *ade*-2—*pdx*—*met*-1.

To work out the recombination frequencies, it is convenient to restructure the table of results to deal with the progeny in two pairwise combinations, *ade*-2 to *pdx*, and *pdx* to *met*-1. This gives the following progeny distributions:

pdx	*ade*-2	155 + 7
+	+	170 + 10
pdx	+	36 +1
+	*ade*-2	30 +2

Which corresponds to 69 recombinants in 411 progeny = 16.8% recombination. In addition:

pdx	+	155 + 36
+	*met*-1	170 + 30
pdx	*met*-1	7 +1
+	+	10 +2

Corresponding to 20 recombinants in 411 progeny, or 4.9% recombination. Thus, the final map can be drawn like this: *ade*-2—16.8—*pdx*—4.9—*met*-1.

There is one further piece of information we can extract from these data. When we first calculated the expected frequency of double recombinant progeny in Section 5.2, we did it by multiplying together the two recombination frequencies with the assumption that recombination events in one interval were independent of recombination events in the other interval. We have here a means of testing that assumption. We know that there were three double recombinant progeny (= 0.73%), and we can calculate that we expect 16.8% of 4.9% double recombinants (= 0.82%). The two clearly do not match. We observe fewer double recombinant progeny than the assumption of randomness in recombination leads us to expect.

This is quite commonly observed. It means that crossing over within one interval is not independent of crossing over within the immediately adjacent interval. Instead, there is some sort of interference between the two crossover events such that the occurrence of one crossover makes a second crossover in the immediate vicinity less likely to occur. This is called *chiasma interference* because it reduces the occurrence of chiasmata (the fancy name for crossovers). What causes it is not known, but the extent of it is measured with a value called the *coefficient of coincidence*, calculated by comparing the expected frequency of double crossing over with the observed frequency:

$$\frac{\text{observed frequency of double crossovers}}{\text{expected frequency of double crossovers}}$$

If crossing over in one interval is fully independent of crossing over in the other, you will observe as many double recombinants as you expect and this coefficient will calculate to the value 1. If there is interference between crossovers, however, such that crossing over in one interval reduces the likelihood of crossing over in the other, then the coefficient will be less than 1. This is what we have found in the *ade*-2—*pdx*—*met*-1 region of the *Coprinus* chromosome. The coefficient of coincidence here is calculated with this equation.

$$\frac{0.0073}{0.0082} = 0.89$$

This corresponds to a mild but positive interference. Calculating the coefficient of coincidence is therefore a means of detecting interference between crossovers. It is usually mildly positive (with coefficients of coincidence in the 0.6–0.9 range) although if the region of chromosome you are working with is located near a recombination hot-spot you may observe more double recombinants than expected, which would give a coefficient of coincidence greater than 1. This is called *negative interference* and has been observed, but only occasionally.

The mapping process we have just described, using two- and three-point crosses, is applicable to all eukaryotes. The only difference is a practical one that arises from the fact that most animals and plants are diploid in their free-living phase. In haploid fungi, progeny genotypes are immediately evident in the phenotypes of the haploid progeny. In diploid organisms, though, testcrosses, in which one of the parents carries the recessive alleles of all of the genes under investigation, must be used to ensure that phenotypes of the diploid progeny directly represent genotype distributions of the gametes that arise from the triply heterozygous parent.

There is one aspect of chromosome mapping that is almost totally limited to fungi, in the sense that only fungi offer a means to perform the analysis as part of their normal life style. This is the ability to map the position of the centromere within its linkage group using progeny segregation data.

5.5 Mapping Centromeres Using Gene Segregations in Tetrads and Eight-Spored Asci (Octads): Single Gene Segregations

A unique feature of many fungi is that it is possible to isolate and analyze all four products of a single meiosis. A few ascomycetes are even more unique because their ascus is so narrow that all division spindles are

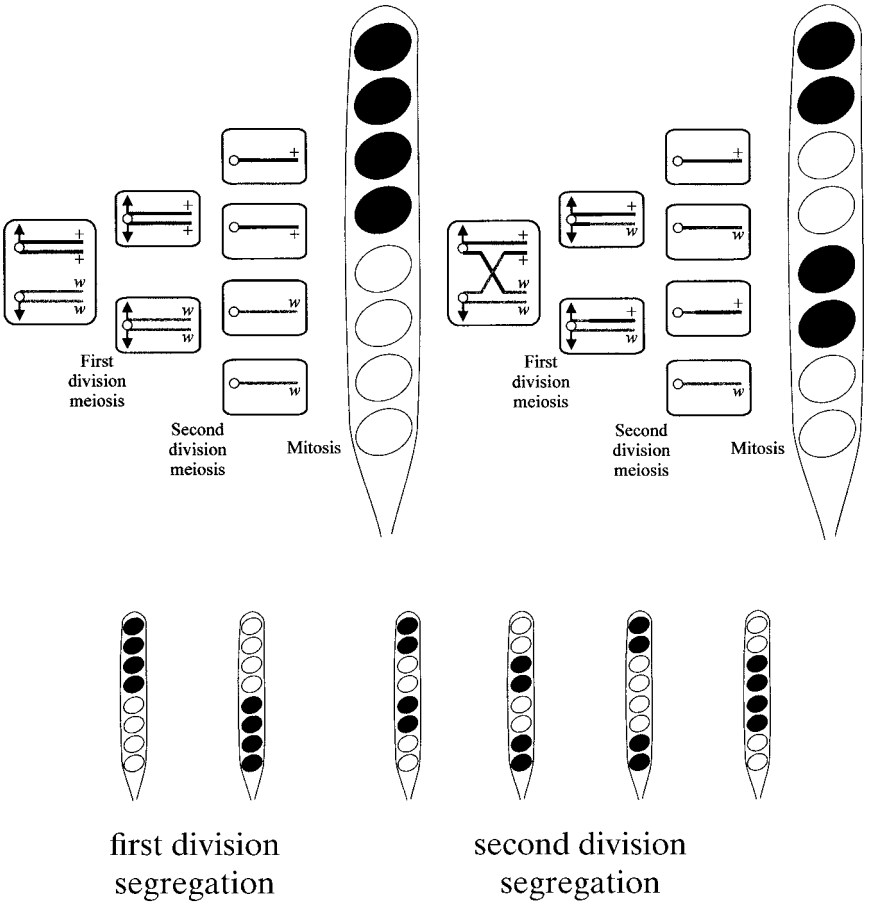


Fig. 5.3. Ascus segregation patterns of ascomycetes like *Sordaria*, *Ascobolus*, and a few species of *Neurospora*, illustrated with the segregation of a spore color phenotype (+ = wild-type black, *w* = white). All division spindles are constrained to the long axis of the ascus. As a result, in these few species the meiotic products come to be arranged in a linear order that strictly depends on the mechanics of meiosis and whether or not a crossover occurs between the gene and its centromere (top panel). Because the ascus has a recognizable top and bottom, and nuclei can rotate before dividing, there are two first-division, and four second-division segregation patterns (bottom panel).

constrained to its long axis. As a result, in these few species the meiotic products are arranged in a linear order that strictly depends on the mechanics of meiosis (Fig. 5.3). This happens in species of *Sordaria*, *Ascobolus*, and a few species of *Neurospora*.

Frequencies of these patterns depend on the frequency of recombination between the gene and its centromere. Thus, by recording the number of second division segregating asci, the distance between the gene and its centromere can be calculated. The convention for calculating recombination frequency (total recombinant progeny as a percentage of total progeny) was unfortunately established before these ascus segregations were identified. Because half the spores in a second division segregation ascus are nonrecombinant, the recombination percentage is equal to the second division segregation percentage divided by two.

As an aside, note that it is not just ascus segregations to which these statements apply; rather, it applies to meiosis in all eukaryotes. Recombination occurs at the four-strand stage of meiosis, but one crossover only ever involves two of the four chromatids. The other two chromatids remain nonrecombinant; therefore, they contribute to the divisor in the recombination frequency equation. As a general rule, therefore, the percentage of meioses in which a crossover occurs equals twice the recombination frequency. A recombination frequency of 20% means that 40% of meioses have a crossover in that interval; additionally, a recombination frequency of 45% corresponds to 90% of meioses that have a crossover between those genes. Not convinced? Well, let's work it out this way: say you have 100 meiotic divisions and 10 of them (= 10%) have a crossover between two particular genes. What is the recombination frequency? One hundred meioses produce 400 chromatids, and each one of which will enter a spore and become a progeny mycelium. In the 10 recombinant meioses the crossover will take place at the four-strand stage, but will involve only two strands in each meiosis, so total recombinant progeny will be $10 \times 2 = 20$. As a result, the recombination frequency calculates to:

$$\frac{20}{400} \times 100\% = 5\%$$

We started out, however, by saying that 10% of the ***meioses*** had a crossover between those two genes . . . it's all a matter of arithmetic.

The following example illustrates in how ascus segregations are used for mapping. Ascospore color in *Sordaria* is controlled by a pair of alleles at one gene locus, the wild-type allele, +, makes spores black and the mutant allele, *w*, results in a white spore. Suppose that a cross was made between a white and a black isolate and the segregation of spore color in 274 asci formed by that cross was as shown in Table 5.4.

You have to be careful reading data presented like this. Each *vertical* column represents "an ascus type," and the cells within each column represent the spores in that ascus type. Because an ascus is a hyphal branch, it has a recognizable bottom (where it is joined to the parental hypha) and a recognizable top (the ascus apex, where the growing tip of the branch used to be). We have numbered each ascospore in Table 5.4, so that ascospore 1

Table 5.4. Progeny octads obtained from a cross in *Sordaria* between the wild type allele, +, which makes the spores black and the mutant allele, *w*, which results in white spores.

	Ascus type (first or second division)					
	2nd	2nd	1st	2nd	1st	2nd
Ascospore 1	+	w	w	+	+	w
Ascospore 2	+	w	w	+	+	w
Ascospore 3	w	+	w	w	+	+
Ascospore 4	w	+	w	w	+	+
Ascospore 5	+	+	+	w	w	w
Ascospore 6	+	+	+	w	w	w
Ascospore 7	w	w	+	+	w	+
Ascospore 8	w	w	+	+	w	+
Totals observed	9	16	129	10	105	5

is at the top and ascospore 8 at the bottom. The extreme right-hand column of Table 5.4 is telling you that in the sample of asci examined, there were five that had two white ascospores at the top, followed by two black, then two white, and finally two black at the bottom. That is a second division segregation pattern. The next column in from the right shows that there were 105 asci that had four black ascospores in the top half of the ascus and four white in the bottom half. This is a first division segregation pattern. Carry on interpreting the columns in the table like that and you will see that both first division segregation patterns are represented (indicated in the heading of the table), and all four possible second division segregation patterns. The data we use for calculation are the totals in the final (bottom) row. These establish that first division segregation asci total $129 + 105 = 234$, and second division segregation asci total $9 + 16 + 10 + 5 = 40$. As a result:

$$\%\ \text{second division segregation} = \frac{40}{274} \times 100 = 14.6\%$$

Remember, in each second division ascus only half the spores derive from chromatids that have undergone crossing over, so the recombination percentage is half of this second division segregation frequency; therefore, the spore color locus shows 7.3% recombination with its centromere.

5.6 Mapping Using Multiple Gene Segregations in Tetrads and Octads

It is also possible to use ascus segregations to study linkage between chromosomal genes. We've called the mutant alleles *a* and *b* in the following examples; so consider two genes with mutant and wild-type alleles *a*/+ and

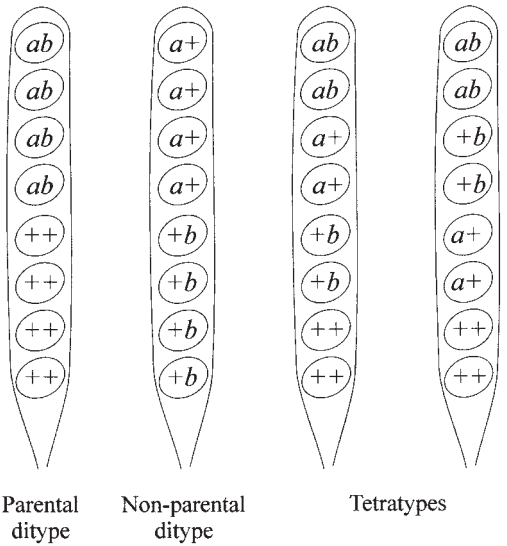


Fig. 5.4. Possible segregation patterns in octads from the cross $ab \times ++$.

b/+. If we make a cross and then analyze spores from individual asci after recording the position of each spore in its ascus, then the possible segregation patterns from the cross $ab \times ++$ are as shown in Fig. 5.4. There is the parental ditype (PD), which has just two types of spore with the parental genotypes; the nonparental ditype (NPD), again with two types of spore, but this time all spores have recombined (or non-parental) genotypes; and the figure shows two forms of tetratype (T), which contain four types of spore, comprising all of the possible genotypes. The relative frequencies of these ascus types depend on whether the genes are linked.

If the genes are unlinked, then: (1) chromosomes assort independently and PD as well as NPD arise with equal frequency; and (2) tetratypes arise when one gene recombines with its centromere. The two sorts shown in Fig. 5.4 show the two alternatives of either gene *a* or gene *b* recombining with its centromere and segregating in the second division of meiosis.

If the genes are linked, then (1) PD arise by chromosome segregation and are common; (2) NPD only arise through a four-strand double crossover and are rare; and (3) tetratypes arise from a single crossover between the genes.

The formula for calculating recombination percentages from this sort of octad analysis is therefore:

$$\%R = \frac{\frac{1}{2}T + NPD}{Total\ Octads} \times 100$$

The number of tetratypes must be divided by two because only two chromatids in a tetratype are recombinant; on the other hand, all spores in NPD asci arise from recombinant chromatids, so they all count as recombinant progeny in this equation.

Tetrad analysis is a very sensitive means of detecting linkage. This is because a PD–NPD ratio of 1:1 is absolutely diagnostic of no linkage. If this ratio is significantly different from 1:1, then linkage is indicated.

For an example of this sort of analysis, we will turn to *Neurospora crassa*, and an experiment in which a strain unable to synthesize riboflavin (symbolized *ribo*; riboflavin is vitamin B_2 and contributes to flavin nucleotide coenzymes) was crossed with a strain unable to synthesize the amino acid tryptophan (*tryp*). The asci that were scored are shown in Table 5.5. Again, each ascus type is represented by the eight spore genotypes in each vertical column, but in this case no distinction has been made between the top and bottom of the ascus (to reduce the number of columns needed to display the results), although we have numbered each column so that we can refer to particular ascus types when required.

We can use these data to construct a chromosome map showing the positions of *ribo*, *tryp*, and their centromere, and we can also extract a bit more information from these data. Taking it one step at a time, first consider the relationship between *ribo* and *tryp*. Ascus types 1, 2, 3, and 4 are all parental ditype, and all of the rest are tetratypes. There are no nonparental ditypes, so clearly the two genes are linked together and presumably they are too close together for any NPD asci to appear in a sample of this size.

Table 5.5. Progeny octads obtained from a cross between a strain of *Neurospora crassa* unable to synthesize riboflavin (*ribo*) and a strain unable to synthesize the amino acid tryptophan (*tryp*).

Ascus type (reference number)									
1	2	3	4	5	6	7	8	9	10
ribo +	+ *tryp*	*ribo* +	*ribo* +	*ribo tryp*	*ribo tryp*	*ribo* +	*ribo* +	*ribo* +	*ribo tryp*
ribo +	+ *tryp*	*ribo* +	*ribo* +	*ribo tryp*	*ribo tryp*	*ribo* +	*ribo* +	*ribo* +	*ribo tryp*
ribo +	*ribo* +	+ *tryp*	+ *tryp*	*ribo* +	*ribo* +	*ribo tryp*	*ribo tryp*	+ +	+ +
ribo +	*ribo* +	+ *tryp*	+ *tryp*	*ribo* +	*ribo* +	*ribo tryp*	*ribo tryp*	+ +	+ +
+ *tryp*	*ribo* +	*ribo* +	+ *tryp*	+ +	+ *tryp*	+ +	+ *tryp*	*ribo tryp*	*ribo* +
+ *tryp*	*ribo* +	*ribo* +	+ *tryp*	+ +	+ *tryp*	+ +	+ *tryp*	*ribo tryp*	*ribo* +
+ *tryp*	+ *tryp*	+ *tryp*	*ribo* +	+ *tryp*	+ +	+ *tryp*	+ +	+ *tryp*	+ *tryp*
+ *tryp*	+ *tryp*	+ *tryp*	*ribo* +	+ *tryp*	+ +	+ *tryp*	+ +	+ *tryp*	+ *tryp*
129	1	2	1	15	13	17	17	2	1

Converting these statements to numbers we have: PD = 129 + 1 + 2 + 1 = 133; NPD = 0; T = 15 + 13 + 17 + 17 + 2 + 1 = 65, total asci scored = 198; so *ribo* and *tryp* are linked and are 16.4 units apart. These are octads, however, so we can determine the position of the centromere by re-interpreting the data in terms of segregation at first or second division of meiosis.

The second step, then, is to examine the relationship between *ribo* and the centromere. We are only interested in second division segregation asci, and looking back at Table 5.5, ascus types 2, 3, 4, 9, and 10 have *ribo* segregating in the second division. In numbers this = (1 + 2 + 1 + 2 + 1) = 7, and 7/198 = 3.5%; therefore, the distance from *ribo* to the centromere is half this, which is 1.8% recombination.

If the same arguments are applied to the relationship between *tryp* and the centromere, second division segregation asci total (1 + 2 + 1 + 15 + 12 + 17 + 17 + 1) = 67, and 67/198 = 33.8%, so it follows that *tryp* shows 16.9% recombination with the centromere. Putting all this together we can draw the map: centromere—1.8—*ribo*—16.4—*tryp*, which is a very nice little map, until somebody asks: "Since when did 1.8 + 16.4 add up to 16.9?"

Indeed, this map illustrates an extremely common feature; recombination in longer intervals tends to be underestimated. These are whole asci, however, and their spore arrangements record all of the events in meiosis. Because they record all events, octads allow us to carry out a complete audit of recombination events. Nothing that follows this should be taken to imply that the preceding approach and analysis is wrong. The asci were categorized correctly and calculations were done properly. In experiments using random spores the analysis could not be taken any further because there are no further data, but these ascus segregations reveal instances of double recombination that would otherwise go undetected.

For example, 67 asci were categorized as second division segregation between *tryp* and the centromere, but refer back to Table 5.5 and you will see that asci numbers 9 and 10 contain some surprises. Ascus 10, for example, has a distribution of spore genotypes that can only be explained as resulting from a three-strand double crossover. This has already been included among the 67 single crossover second division segregation asci, but it represents two crossovers, so we must add one for the second crossover; and total recombination events between *tryp* and the centromere so far total 67 + 1 = 68.

The two octads in Category 9 have the first division segregation pattern for *tryp* and, for that reason, have not been included in the calculation of recombination between *tryp* and its centromere at all yet. These asci, however, do show second division segregation for *ribo*, and have been included in that calculation. Think about that fact, however: *ribo* is closer to the centromere than is *tryp*, and we know that the second division segregation for *ribo* means that these asci have a crossover between *ribo* and the centromere. The only way that the outer gene can segregate in the first

division when the inner segregates in the second division is if a second crossover takes place between the two genes. Thus, these two asci each arise from two-strand double crossovers between *tryp* and its centromere. A total of four crossovers have not been included yet, and these must be added in to bring the overall total count of recombination events between *tryp* and the centromere to 68 + 4 = 72.

This allows corrections of the recombination values as follows:

$$\text{Between } tryp \text{ and centromere, recombination percentage} = \frac{\frac{1}{2}72}{198} \times 100 = 18.182\%.$$

$$\text{Between } ribo \text{ and } tryp, \text{ recombination percentage} = \frac{\frac{1}{2}65}{198} \times 100 = 16.414\%.$$

$$\text{Between } ribo \text{ and centromere, recombination percentage} = \frac{\frac{1}{2}7}{198} \times 100 = 1.768\%.$$

This gives us perfect additivity: 1.768 + 16.414 = 18.182. The message of this rather heavy-handed analysis is that the commonly encountered underestimation of recombination percentage over longer intervals is due to the occurrence of double crossovers. The second crossover reverses the genotype change caused by the first with the result that the fact that the progeny are recombinant is no longer apparent.

Analysis of spore segregations in asci has revealed a great deal about eukaryotic genetic mechanisms. An immediately relevant example is that octad analysis enables the distribution of crossovers to be established. The spore patterns that arise from two-strand double, the two forms of three-strand double, and four-strand double recombinants can all be distinguished and counted (Fig. 5.5). All of these are found to occur and they do occur at equal frequency. This means that there is no limitation in the choice of chromatids by successive crossovers (i.e., there is no chromatid interference).

Furthermore, although the six patterns shown in Fig. 5.3 are normally expected from crosses between alleles of the same gene (e.g., black × white spore character), in rare asci aberrant segregations including 5 black : 3 white are observed. An explanation of how such octads can arise reveals some details of the mechanism of recombination, which we will discuss in Chapter 6.

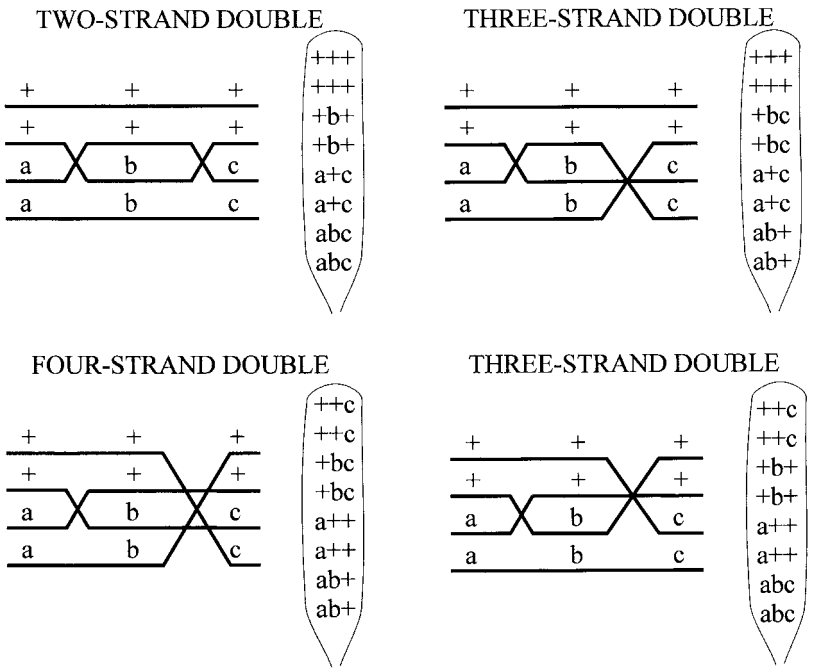


Fig. 5.5. Line diagrams of the four different types of double crossover events, shown here as occurring at the four-strand stage (prophase of meiosis) in a region of the chromosome heterozygous for mutant genes *a*, *b*, and *c*. The lists on the right show the genotypes of ascospores, as they would appear in ascus octads following these recombination events. Each pattern is distinct, so all of the double crossover events can be identified and their frequencies of occurrence determined.

5.7 Unordered Tetrads

Isolating the spores from the ascus one by one in order, noting their original positions, and then testing their colonies after germination for whatever phenotypic characters are segregating in the cross quite readily establish ascus spore segregation patterns. The initial ascospore isolation needs to be done with the aid of a microscope, and may require a micromanipulator, so the technique can be tedious, but it is not especially difficult. Since the order of the four chromatids generated by the meiotic division is maintained by the shape and size of the ascus (even after the postmeiotic mitosis), these are called *ordered octads*. As we have noted, asci narrow enough to maintain the inherent order of meiosis like this are produced by very few species. Many, many other species produce "fatter" asci in which the spores are jumbled up. Then think of all those basidiomycetes in which meiotic products are packaged into basidiospores that are neatly arranged in sets of four on top of the basidia; there is no clue left here of any inherent meiotic order.

These are called *unordered tetrads* because the meiotic order has been lost, yet the products of individual meioses in each of these cases are kept together and all that has been said earlier about segregation in meiosis still, basically, applies.

The analysis of the basidiospore tetrads of *Coprinus*, for example, falls into four consecutive procedures: (1) tetrads must be isolated from fruit body tissues and the spores separated from one another; (2) the spores must be germinated; (3) the sporelings are grown up into colonies; and (4) the colonies are tested. Basidiospore tetrads can be picked off from fresh fragments of the gills with a glass needle, relying on electrostatic attraction for their physical removal. The needle is a thin filament of ordinary white soda glass with a small ball (about 40 μm diameter) on the end. It is necessary to use a micromanipulator because of the small size of the tetrads (approximately 10 μm square) and their high numerical density on the gill surface (approximately 3000 mm^{-2}).

Immediately after removal from the gill, the tetrad is placed onto an agar surface and the four spores separated from one another with the micromanipulator needle. All of these steps need to be done with the aid of a binocular dissecting microscope. There are two main difficulties inherent in this method of picking-off directly from the gill surface. The high numerical density of tetrads makes it very difficult to be sure that no extraneous spores have been picked-off until the tetrads are parked on the agar (by which time it is too late!). Also, it is impossible to distinguish unripe from ripe tetrads, so many abortive attempts to pick up tetrads might be made before one is successfully removed.

In addition to these difficulties there are problems arising from poor spore germination and relatively high losses to contamination. It is difficult to isolate large numbers of tetrads because of such technical difficulties. Another point is that because the members of a tetrad do not represent independent events, the analysis of 100 tetrads (quite a formidable undertaking in itself) is not equivalent to the analysis of 400 random spores. In fact, on average, one tetrad contributes about as much data as two random spores. Thus, in comparison with random spore analyses, tetrad analysis suffers from the effects of small sample sizes. Consequently, when mapping from tetrad analysis data, the recombination values obtained are usually only approximate; however, it is still the only way of mapping centromeres from segregation data.

The method of analysis used to extract the required data from unordered tetrad segregations is slightly different from that used with ordered tetrads. It is an indirect method in that it is necessary to have other genes segregating in the same tetrad as the gene under test. There is a requirement for one other unlinked gene that is known to be close to its centromere so that it nearly always segregates at the first division (known as a *centromere marker*). On the other hand, at least two other genes are required in the cross, neither of which need be known centromere markers, but at least one

of which must be independent of the gene whose centromere distance is to be determined.

The mathematical theory that underlyies the approach depends on the relative frequencies of the three different types of tetrad obtainable from a cross-heterozygous at two loci, which was discussed at the beginning of Section 5.6: the parental ditype (PD), the nonparental ditype (NPD), and the tetratype (T). The order of the spores in the tetrad is irrelevant to identifying these three categories. If there are four spore genotypes in the set, it must be a tetratype; if there are only two spore genotypes, it must be one of the ditypes. Comparing the spore genotypes with the parent genotypes will show you which ditype; however, different frequencies of these tetrad types can occur depending on the relative positions of the genes to each other and to their centromeres. As stated earlier, if the two genes are linked then the proportion of tetratypes is dependent on the frequency of crossing over in the interval between them, and with linked genes nonparental ditypes arise only from four-strand double exchanges between them.

If the two genes are on different chromosomes, then the tetratype frequency is a reflection of the rate of crossing over between the genes and their respective centromeres. If both genes are very close to their centromeres, tetratypes will be rare. Tetratype frequencies become greater as genes with larger centromere distances are used. Indeed, the rules are that tetratype tetrads arise in all cases where the genes individually show second division segregation and in half of the cases where both segregate at the second division simultaneously. If we can convert that into a mathematical statement we could use it for analyzing crosses.

Let us say we have two unlinked genes, *a* and *b*, that have second division segregation frequencies $\boldsymbol{a}$ and $\boldsymbol{b}$ respectively. It is fairly clear from the diagrams in Fig. 5.3 that a tetratype tetrad will be produced by a crossover between *a* and its centromere when *b* remains unrecombined. Now, the frequency of *b* that remains unrecombined is given by $(1 - \boldsymbol{b})$, and a proportion *a* of those will have a crossover between *a* and its centromere, so the frequency of tetratypes due to recombination of *a* alone = $\boldsymbol{a}(1 - \boldsymbol{b})$. By a similar line of argument, the frequency of tetratypes due to recombination of *b* alone = $\boldsymbol{b}(1 - \boldsymbol{a})$. The proportion of meioses in which both genes recombine with their centromere is given by multiplying their second division segregation frequencies together = $\boldsymbol{ab}$. Because of the independent segregation of daughter centromeres in the second meiotic division, however, these meioses give rise to a ratio of 1PD:2T:1NPD tetrads. The frequency of tetratypes arising when both genes segregate at the second division simultaneously therefore = $\boldsymbol{ab}/2$.

The overall frequency of tetratypes is the sum of these three cases = $\boldsymbol{a}(1 - \boldsymbol{b}) + \boldsymbol{b}(1 - \boldsymbol{a}) + \boldsymbol{ab}/2$, which resolves to the expression:

$$\text{tetratype frequency} = \boldsymbol{a} + \boldsymbol{b} - \frac{3\boldsymbol{ab}}{2}$$

If you have previous knowledge of either ***a*** or ***b***, it is possible to determine the other directly using this equation. As a result, if you have one reliable centromere marker, you can use it in this way to establish the centromere distance of any other unlinked gene.

All is not lost, however, if both ***a*** and ***b*** are unknown. By introducing a third independent gene into the cross the required second division segregation frequencies can be calculated without prior knowledge of any of them. With a third pair of alleles, and considering the genes two at a time, three equations of the form shown above are obtained, which then can be resolved into three further equations, each of which gives one of the second division segregation frequencies.

To illustrate, we will add gene *c* (with a second division segregation frequency ***c***) to the *a* and *b* loci just considered. With *c* in the cross, there are three pairwise combinations, so we can write three equations for the tetratype frequency (which we will call T_{ab}, T_{ac}, and T_{bc}): Solutions of these three simultaneous equations for ***a***, ***b***, and ***c*** are as follows:

$$T_{ab} = \boldsymbol{a} + \boldsymbol{b} - \frac{3\boldsymbol{ab}}{2}, \qquad T_{ac} = \boldsymbol{a} + \boldsymbol{c} - \frac{3\boldsymbol{ac}}{2}, \qquad T_{bc} = \boldsymbol{b} + \boldsymbol{c} - \frac{3\boldsymbol{bc}}{2}$$

$$\boldsymbol{a} = \frac{2}{3}\left[1 \pm \sqrt{\frac{4 - 6T_{ab} - 6T_{ac} + 9T_{ab}T_{ac}}{4 - 6T_{bc}}}\right],$$

$$\boldsymbol{b} = \frac{2}{3}\left[1 \pm \sqrt{\frac{4 - 6T_{ab} - 6T_{bc} + 9T_{ab}T_{bc}}{4 - 6T_{ac}}}\right],$$

$$\boldsymbol{c} = \frac{2}{3}\left[1 \pm \sqrt{\frac{4 - 6T_{bc} - 6T_{ac} + 9T_{bc}T_{ac}}{4 - 6T_{ab}}}\right],$$

These equations look worse than they really are; the arithmetic is essentially straightforward, and the three tetratype frequencies are all that's required for the analysis. There are two solutions to each of these equations. Negative ones are obviously unreal, and any positive value greater than 0.67 is unlikely because second division values greater than 67% are extremely rare.

For our worked example, we turn back to *Coprinus cinereus*, and crosses between a wild type and an auxotroph unable to synthesize nicotinic acid, called *nic*-4, in which the mating type of all progeny was scored so that the *A* and *B* mating-type factors made up the total of three genes. A total of 336 tetrads were analyzed. They were scored in a number of batches, so before doing the overall analysis it was necessary to verify the homogeneity of the data using contingency tables of the tetrad types between each gene pair observed in each batch. The χ^2 values obtained corresponded in each case to a probability of about 20% that the data were satisfactorily homogeneous. The data were correspondingly bulked into Table 5.6.

Table 5.6. A bulked progeny sample of 336 unordered tetrads of *Coprinus cinereus* scored for the auxotroph *nic*-4, and the *A* and *B* mating-type factors.

	Tetrad types			
Gene pair	PD	NPD	T	%Tetratypes
A/*B*	100	99	137	40.8
A/*nic*-4	92	99	145	43.2
nic-4/*B*	136	132	68	20.2

Testing the three PD–NPD ratios for deviation from 1:1 gave χ^2 values of 0.005 (*A*/*B*), 0.26 (*A*/*nic*-4), and 0.06 (*nic*-4/*B*). With one degree of freedom, these χ^2 values correspond to very high probabilities that deviations from 1:1 were due to chance variation. As a result, the PD–NPD ratios provide good evidence that the three genes are unlinked, and we can continue with the analysis.

The second division segregation frequency (SDSF) of mating-type factor *A* is given by this equation:

$$= \frac{2}{3}\left[1 \pm \sqrt{\frac{4-(6\times 0.408)-(6\times 0.432)+(9\times 0.408\times 0.432)}{4-(6\times 0.202)}}\right]$$

$$= \frac{2}{3}\left[1 \pm \sqrt{\frac{0.546}{2.788}}\right] = \frac{2}{3}[1 \pm \sqrt{0.196}] = \frac{2}{3}[1 \pm 0.443]$$

The two solutions of this equation are 0.962 or 0.372. The former is unreal (96% second division segregation?), which means that the mating-type factor *A* shows 37.2% second division segregation (equivalent to a recombination frequency of 18.6%) with its centromere. If we continue to the other two genes in the cross, we have:

SDSF of mating factor *B*

$$= \frac{2}{3}\left[1 \pm \sqrt{\frac{4-(6\times 0.408)-(6\times 0.202)+(9\times 0.408\times 0.202)}{4-(6\times 0.432)}}\right]$$

$$= \frac{2}{3}\left[1 \pm \sqrt{\frac{1.082}{1.408}}\right] = \frac{2}{3}[1 \pm \sqrt{0.768}] = \frac{2}{3}[1 \pm 0.876] = 1.25 \text{ or } 0.083.$$

and

$$\text{SDSF of } nic\text{-}4 = \frac{2}{3}\left[1 \pm \sqrt{\frac{4-(6\times 0.202)-(6\times 0.432)+(9\times 0.202\times 0.432)}{4-(6\times 0.408)}}\right]$$

$$= \frac{2}{3}\left[1 \pm \sqrt{\frac{0.981}{1.552}}\right] = \frac{2}{3}[1 \pm \sqrt{0.632}] = \frac{2}{3}[1 \pm 0.795] = 1.20 \text{ or } 0.137.$$

Second division segregation frequencies in excess of 100% are obviously unreal, so we conclude that mating-type factor *B* shows 8.3% second division segregation (4.2% recombination) with its centromere, whereas the equivalent figures for *nic*-4 are 13.7% SDSF and 6.9% recombination.

This analysis reveals the very useful fact that mating-type factor *B* in *Coprinus cinereus* is sufficiently close to its centromere for multiple recombination events to be rare. Since all crosses require heterozygosity in the mating-type factors, and it is relatively simple to carry out mating-type testing on the progeny, the B factor is a convenient centromere marker, which is "built-in" to all crosses in *C. cinereus*.

5.8 Linkage Analysis to Linkage Map

With a combination of random spore and tetrad analyses, a good picture of the chromosomal locations of a wide range of functional genes can be established. Example chromosome maps for a few fungi can be found in several of the references we quote at the end of this chapter. The resolution and detail of maps constructed using genetic segregations is limited by the ability to analyze sufficient progeny to detect rare crossovers. The most detailed fungal segregation map is that of *Saccharomyces cerevisiae*, which shows the locations of approximately 1200 genetic markers. This is an average of about one for each 10 kbp of DNA. Some discrepancies became evident when the complete sequence of chromosome III of *S. cerevisiae* became available, but the discrepancies are very few in number, so for the most part segregation analysis gives an accurate view of gene arrangements.

Linkage analysis in *S. cerevisiae* is very efficient, and it is unlikely that the resolution of the linkage map achieved in this yeast will be equaled in other fungi. Indeed, in many cases the level of detail attained in yeast cannot even be approached in other fungi. The inaccuracies noted in the yeast linkage map when compared with the DNA sequence suggests that linkage maps cannot be expected to provide a detailed physical map of the DNA. Rather, the linkage map is a representation of the behavior of genes and chromosomes as they progress through the meiotic division cycle. This has enormous practical significance for our understanding of natural populations, prediction of progeny populations, and development of altered genotypes using classic genetic approaches.

The discussion so far has concentrated on functional genes, but much the same analyses using genetic segregations can be employed to construct gene maps showing the chromosomal locations of DNA markers such as restriction fragment length polymorphisms. Such markers, however, also lend themselves to alternative mapping techniques that depend on direct analysis of nucleic acid molecules, and we will discuss this in Section 8.2.

5.9 Tetrad Segregations Leading to Secondary Homothallism

Some organisms use the mechanics of meiosis to control their sexuality. By producing fewer spores than there are nuclei to accommodate, nuclear migration after meiosis can make the spores homothallic. *Neurospora tetrasperma* is an example. Unlike its relative *N. crassa*, *N. tetrasperma* is homothallic, which means that the mycelium that grows from a single ascospore is self-fertile and is able to produces mature perithecia. The species shows secondary homothallism; this differs from the primary homothallism of *Aspergillus nidulans* and *Sordaria macrospora* because *N. tetrasperma* has a mating-type locus that exists in two forms, A and a, and ascospores contain two nuclei, there being one of each mating type. The mating-type locus is so close to its centromere that recombination between the two is extremely rare. As a result, the mating-type locus almost always segregates at the first division of meiosis. During the second meiotic division the division spindles are normally parallel, or at least overlap, and the subsequent mitotic division spindles are more or less at right angles to the long axis of the ascus. Ascospore walls surround one nucleus of each type. Thus, on germination, a single ascospore gives rise to a heterokaryotic mycelium that is heterozygous for mating type, and therefore fertile.

The most important instance of secondary homothallism probably occurs in the cultivated mushroom, *Agaricus bisporus*, which, as the specific name implies, usually forms only two spores on the basidium. Each spore must, consequently, be provided with two meiotic products. As the diploid nucleus prior to meiosis was heterozygous at the mating-type factor, there are two progeny nuclei of each mating type in the basidium prior to spore formation. As a result, three different genotypes are possible in the spores according to which nuclei migrate together: $(A + A)$ and $(a + a)$, both of which would germinate to produce sterile homokaryons, and $(A + a)$ which would germinate into a fertile heterokaryon. This is secondary homothallism again because this single progeny spore germinates to give rise to a fertile mycelium.

Random segregation of nuclei would give a ratio of 2:1 for heterokaryotic to homokaryotic progeny in any such secondarily homothallic species that packs two compatible nuclei into the same spore. In field collected isolates of *Agaricus bisporus*, however, a significant deviation from this ratio is observed, and the heterokaryon is favored. It seems, therefore, that the nuclear migration mechanism may be able to sort nuclei as well as transport them. There are *Coprinus* species that have two-spored basidia, and the spores may have nuclei of compatible mating types, which germinate to give fertile dikaryons in a way analogous to secondarily homothallic *N. tetrasperma* and *A. bisporus*. It has been possible to separate the nuclei from the homothallic dikaryons into monokaryons that were then paired experimentally to study the mating-type specificities. Several different *A* and *B* factors were discovered, which suggests that the secondarily homothallic

condition originated from a heterothallic one several times in the evolution of the species.

5.10 Gene Segregation During the Mitotic Division Cycle

In the middle of the twentieth century, it became evident that meiotic segregations were not the only way of making maps of chromosomes. Mitotic segregations can also be analyzed and are a convenient way of mapping chromosomes. The approach is applicable to any fungus that is normally haploid, although the first step is the selection of diploids that arise spontaneously through nuclear fusion at a rate of about one in every 10^6 or 10^7 mitoses.

The pioneering work was done with the ordinarily haploid filamentous fungus *Aspergillus nidulans*. Selection of diploid strains is a little easier in *Aspergillus* because its conidia are always uninucleate. As a result rare diploid conidia can be selected from among a large spore population obtained from a heterokaryon by selecting for a heterozygous phenotype. Uninucleate spores cannot be heterokaryotic, so conidia that express a heterozygous phenotype must contain both homologues of at least one pair of chromosomes, and may be completely diploid. If you make a heterokaryon between two recessive auxotrophs, therefore, you would expect that diploid spores would be the only conidia able to grow on minimal medium. Diploid conidia are larger than haploid conidia, with about twice the volume; they also, of course, contain twice the haploid amount of DNA.

This sort of nutritional selection is an automatic method that certainly works efficiently, but it limits the number of nutritional markers that can be used in any experimental crosses. An especially useful feature of *A. nidulans* (not true for all fungi), however, is that the color of the conidium depends on its own genotype. A heterokaryon made between two non-allelic, recessive, spore color mutants, say white-spored and yellow-spored strains, will consequently produce large numbers of haploid white and yellow conidia together with very occasional sectors of diploid conidia with the wild type dark green color. Using color selection leaves open the possibility of having several (unselected) nutritional markers in the cross, but it requires close scrutiny of the cultures.

Nutritional selective methods have been used to isolate diploids from many normally haploid fungi. This includes basidiomycetes such as *Schizophyllum commune* and the agaric *Coprinus cinereus*, but it is especially important in those fungi in which the known life cycle lacks sexual reproduction. This group includes several commercially important species, including *Aspergillus niger*, *A. oryzae*, *A. flavus*, *Penicillium chrysogenum*, and plant pathogens like *P. expansum* and *P. digitatum*.

Diploids are generally sufficiently stable to grow into diploid vegetative colonies, but these do produce rare sectors that show segregation of the

originally heterozygous genes. This type of segregation, which is also based on mitotic recombination, can be used for genetic mapping. The method was developed first with *Aspergillus nidulans*, in which segregant sectors of the mycelium could be recognized by the color of their spores. In *A. nidulans*, mitotic crossing over has a frequency of about two per thousand mitotic divisions and haploidization about one per thousand mitotic divisions. In a reversal of the procedure used for identifying diploid sectors, segregants can be identified by the appearance of yellow- or white-spored sectors against the background of dark green spores of a parental diploid colony heterozygous for recessive color mutations.

Other methods for selecting segregants are needed where spore color cannot be used (i.e., monitoring differential growth between faster-growing haploid sectors and slower-growing diploid sectors). Partially dominant inhibitor-resistant mutants have been used, but suppressors of auxotrophic mutants provide the clearest example. A diploid homozygous for an auxotroph and heterozygous for the (recessive) suppressor cannot itself grow on minimal medium, but it can segregate haploids (or homozygous partial diploids) able to grow on minimal medium. These mitotic segregants from the diploid prove to be haploid (i.e., produced by a process of regular chromosome loss during successive aberrant mitoses called *haploidization*), partial diploids (i.e., aneuploids stabilized during the chromosome loss sequence), or diploids showing segregation for a few linked genetic markers, and remaining heterozygous for the others. Haploidization is caused by nondisjunction (i.e., improper transport of chromosomes to the poles of the division spindle during mitosis) that result in random chromosome loss over several divisions. Thus, the diploid is reduced to a haploid state through a series of aneuploid intermediates.

We will examine some results from a typical experiment with *Aspergillus nidulans* in which the original diploid was heterozygous for both white (*w*) and yellow (*y*) conidia. These two genes are on different chromosomes, and the *y* chromosome also carried auxotrophic mutations called *ade* (adenine requirement), *pro* (proline requirement), *paba* (requirement for the vitamin *p*-aminobenzoic acid), and *bio* (requirement for the vitamin biotin). The parental diploid had the chromosomal constitution shown in Fig. 5.6a.

Segregants from this diploid were identified on the basis of spore color: the parental diploid produces dark green conidia, but sectors with yellow spores and sectors with white spores are occasionally found. Among the yellow-spored segregants (Fig. 5.6) were strains that were: (1) prototrophic diploids homozygous for yellow, caused by a crossover between *paba* and *y* (Fig. 5.6b); (2) two sorts of yellow diploid homozygous for *paba*, and therefore auxotrophic for *p*-aminobenzoic acid, one caused by a crossover between *bio* and *paba* in the parental diploid, whereas the other was caused by a crossover between *pro* and the centromere (Fig. 5.6c); (3) haploids caused by haploidization without crossing over (Fig. 5.6d); and (4) haploids caused by haploidization after a crossover, in this case, a crossover somewhere between *paba* and the centromere (Fig. 5.6e).

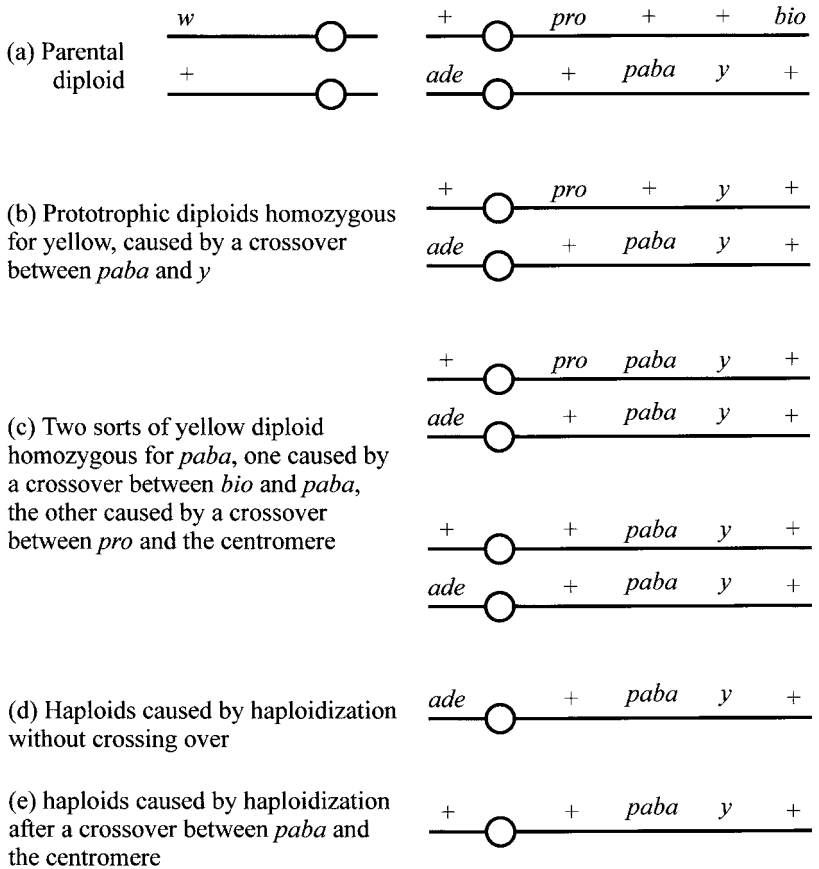


Fig. 5.6. Yellow-spored mitotic segregants obtained from an experiment with *Aspergillus nidulans* in which the original diploid was heterozygous for both white (*w*) and yellow (*y*) conidia. The parental diploid had the chromosomal constitution shown in (a); the crossovers referred to in the other sections of the figure occurred in one chromosome of this genotype before mitotic segregation.

White-spored segregants resulted from homozygosity or haploidization of *w* (Fig. 5.7). White-spored genotypes observed were: (5) prototrophic white-spored diploids caused by a crossover between *w* and the centromere (Fig 5.7b); and (6) two sorts of white-spored haploids caused by haploidization without crossing over (Fig 5.7c). White-spored haploid segregants that require proline and biotin were observed in about the same frequency as those that require adenine and *p*-aminobenzoic acid, which shows that the chromosome carrying these auxotrophic markers segregated independently of the white chromosome during the haploidization process.

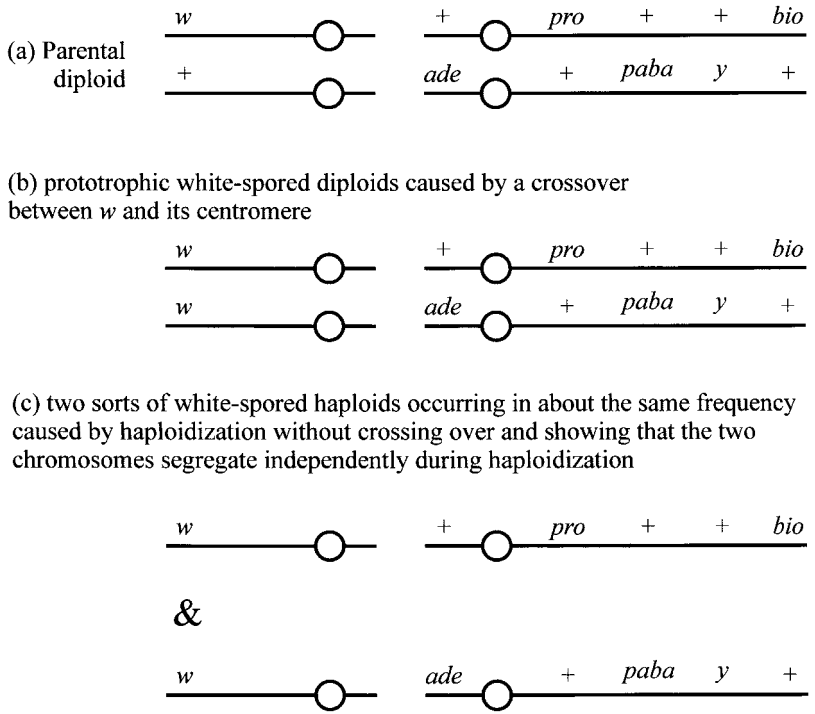


Fig. 5.7. White-spored mitotic segregants obtained from an experiment with *Aspergillus nidulans* in which the original diploid was heterozygous for both white (*w*) and yellow (*y*) conidia. The parental diploid had the chromosomal constitution shown in (a), and the crossovers referred to in the other sections of the figure occurred in one chromosome of this genotype before mitotic segregation.

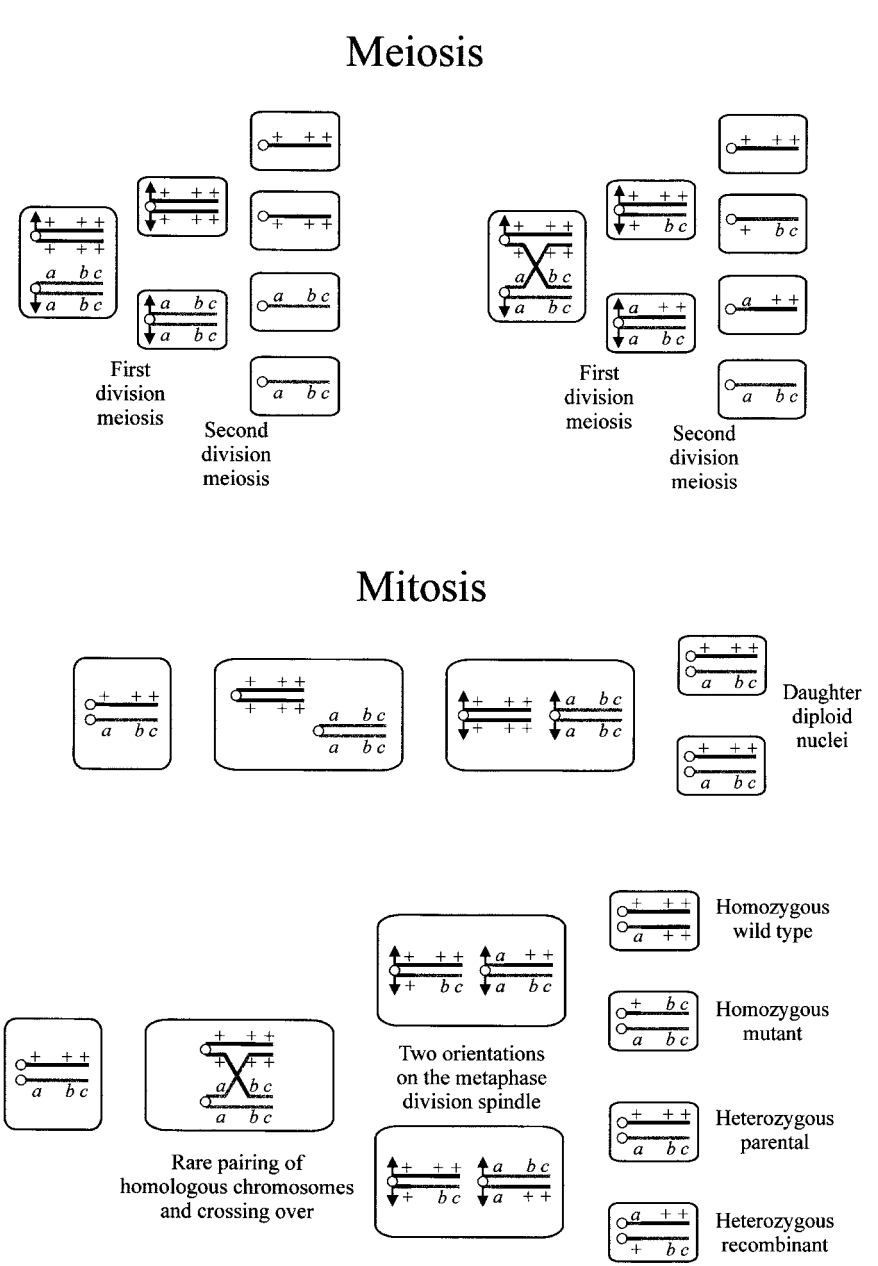
The yellow diploid segregants show that mitotic crossing over is a reality, so let's see what lessons can be learned from this example. The key to understanding is to remember the crucial differences in chromosome behavior during meiosis and mitosis. In meiosis, homologous chromosomes undergo synaptic pairing, and they take their place on the first division spindle as a bivalent comprising two chromosomes, each divided into two chromatids (this is the four-strand stage as illustrated in Figs 5.2, 5.3, and 5.5). At the first division of meiosis, the so far undivided homologous (that is, maternal and paternal) centromeres are taken to opposite poles of the division spindle (Fig. 5.2). *None of this happens in mitosis.*

In mitosis homologous chromosomes do not line up with one another, so there is no synapsis. The fact that recombinant diploid segregants can be obtained from mitotic crossing over indicates that occasional exchanges occur between homologous chromosomes during mitosis. Data obtained from other experiments have demonstrated that mitotic crossing over is a reciprocal event, that is, the recombination results in two homologous

chromosomes with reciprocally recombinant arms (Fig. 5.8), but it is clearly extremely rare. In fact, mitotic crossovers are too rare for double exchanges to be a problem in genetic analysis. Mitotic crossing over can be visualized as very similar to meiotic crossing over, but the consequences in terms of the genotypes of progeny nuclei differs because chromosome segregation differs between meiosis and mitosis. If a crossover takes place in mitosis the two chromosomes involved do not stay together (as they do in meiosis); rather, they separate and reach the equator of the division spindle independently. The two (homologous) chromosomes, which are, of course, divided into two daughter chromatids, then behave independently. In meiosis, the two reciprocally recombinant chromatids *must* end up in different haploid daughter nuclei (Fig. 5.8).

Mitosis, however, produces diploid daughter nuclei by sending one daughter chromatid of each homologous chromosome to each pole of the division spindle. As long as the rule that daughter chromatids must go to opposite poles of the division spindle is upheld, there is no other constraint. The daughter chromatids of a pair of homologous chromosomes segregate independently of each other. As a result, following a mitotic crossover, the two reciprocally recombinant chromatids may pass to opposite poles at the subsequent anaphase stage of mitosis; alternatively, and with equal probability, they may go to the same pole. The importance of this is that in the former case (where a recombinant chromatid is accompanied by a parental chromatid) all markers between the crossover and the end of the chromosome will become homozygous (Fig. 5.8).

For practical analysis, the rule is that in diploid segregants homozygosis for one gene will always be accompanied by homozygosity of any genes distal to it in the same chromosome arm, but not necessarily by homozygosity of genes more proximal to it. This general characteristic of mitotic segregation is what allows us to deduce the genetic map of the segment of chromosome involved, and is illustrated in the yellow diploid segregants in Fig. 5.6. Homozygosity for *y* was invariably accompanied by homozygosity for the wild-type allele of *bio*. This fact places the *bio* gene distal to *y* in the same arm of the chromosome. Homozygosity for *paba* was similarly frequent but not invariable in diploids homozygous for *y*, whereas homozygosity of *pro* was even less frequent in the yellow homozygotes. This pattern indicates that *paba* and *pro* were proximal to *y* in the same arm; that is, on the centromere side of *y*, with *pro* the nearer to the centromere. Noting that co-segregation of *ade* and *paba* in yellow haploids suggests that the two genes are linked, the absence of yellow diploid segregants homozygous for *ade* allows us to infer that this gene may be in the other arm of the chromosome (i.e., other data confirm that homozygosis of one arm of a chromosome occurs independently of homozygosity of the other arm). The frequencies of the different sorts of homozygous diploid segregants are a measure of relative map distance between the genes. But it must be stressed that mitotic recombination is rare, so it is not easy to assemble a sample of independent segregants large enough to make frequency measurements


Meiosis
First division meiosis
Second division meiosis
First division meiosis
Second division meiosis
Mitosis
Daughter diploid nuclei
Rare pairing of homologous chromosomes and crossing over
Two orientations on the metaphase division spindle
Homozygous wild type
Homozygous mutant
Heterozygous parental
Heterozygous recombinant

reliable. Nevertheless, even a few segregants can give absolute guidance about gene order relative to the end of the chromosome arm.

In most organisms, the analysis of meiotic products is usually the easiest way of mapping the genome; however, use of mitotic segregations for genetic analysis does offer some worthwhile advantages over meiotic analysis to the experimenter. Relatively few segregants can provide a considerable amount of information about relative positions of genes on the chromosome. Even one diploid segregant in which linked genes *a* and *b* become homozygous simultaneously provides evidence that they are in the same chromosome arm, and a second segregant in which *a* becomes homozygous alone shows almost certainly that *a* is distal and *b* proximal with respect to the centromere.

Meiotic and mitotic linkage maps show the same gene orders, but the spacing between the genes differs, which implies different distributions of crossovers in mitosis and meiosis. The overwhelming advantage of mitotic analysis, though, is in the formation of haploids. Because mitotic recombination is so rare, genes on the same chromosome show complete linkage during haploidization. Genes reassort freely if they are not on the same chromosome. Thus, linkage group assignments can be made far more efficiently than is possible in meiotic analysis.

Application of the approach to other organisms depends on the stability of the diploid state. This varies enormously between species. *Verticillium albo-atrum* seems to hold the record for instability. A random sample of

Fig. 5.8. Comparison of the segregation mechanisms of meiosis and mitosis. We show meiosis at the top, on the left without recombination, on the right incorporating a crossover event. These are the perfectly normal segregation processes that are also illustrated in Fig. 5.3. Mitosis, with and without a crossover, is shown in the bottom part of the figure. Mitosis starts with replication of the parental chromosomes, but the two homologues do not normally associate with one another. The replicated chromosomes subsequently align on the mitotic division spindle independently, and the rule is that one daughter chromatid of each replicated chromosome segregates into each daughter nucleus, so that the two daughter diploid nuclei have the same genotype as the parent. The replicated chromosomes do very rarely associate with one another sufficiently closely for a crossover to occur. This results in two of the daughter chromatids being recombinant. However, the recombinant replicated chromosomes still independently align on the mitotic division spindle, and they still follow the rule that one daughter chromatid of each replicated chromosome segregates into each daughter nucleus. Because of the crossover, there are two possible orientations on the mitotic division spindle. One produces two diploid daughter nuclei that are homozygous from the point of the crossover to the end of the chromosome arm (i.e., one homozygous mutant; one homozygous wild type). The alternative orientation produces two heterozygous diploid daughter nuclei, but one contains the two parental chromatids, and is exactly the same as the parental nucleus, whereas the other receives the two recombinant chromatids.

540 conidia gave 37 segregants. *Coprinus radiatus* diploids are also unstable, although the closely related *C. cinereus* has segregant frequencies of around 10^{-4}. Diploid segregants due to mitotic crossing over occur with similar frequency (range 10^{-3}–10^{-4}) in *Saccharomyces cerevisiae* in which diploidy is a normal part of the life cycle. Diploids of *Ustilago maydis* seem to be completely stable, but mitotic crossing over is inducible by ultraviolet light. X-rays, nitrous acid, mitomycin C, fluorodeoxyuridine, and the amino acid analogue *p*-fluorophenylalanine have also been used to increase frequencies of segregation from diploids. The common link between these agents seems to be that they cause damage to DNA that induces repair mechanisms that themselves initiate processes leading to mitotic crossing over and haploidization.

We have described a number of separate events that occur during vegetative fungal growth that might be arranged into a sequence. The fusion of genetically different haploid nuclei in a heterokaryon followed by mitotic crossing over and completed by haploidization, is a sequence termed the *parasexual cycle*. On the face of it, the parasexual cycle has much the same effect as the sexual cycle by reassorting and recombining genes, thereby increasing genetic variation within the species. A plausible argument can be made that the parasexual cycle could be an alternative to sex in imperfect fungi, but there is not much clear evidence for this. Indeed, not a great deal of practical use has been made of the parasexual cycle in the laboratory, even though several commercial processes depend on imperfect fungi like *Penicillium chrysogenum*. It is ironic that the technique has found its most extensive application in human genetics. A very large proportion of the gene assignments to human chromosomes were made using the analogous cycle: mouse + human cell forming a hybrid fusion cell that suffers successive loss of chromosomes during subsequent mitoses. Aneuploid cell lines, which are sufficiently stable for genetic and cytogenetic characterization, are eventually formed, and co-segregation of genes reveals linkage.

5.11 Cytoplasmic Segregations: Mitochondria, Plasmids, Viruses, and Prions

Our focus so far has been the segregation of nuclear genes, but there are a number of cytoplasmic genetic elements that affect the fungal phenotype and which depend upon some of the features of the sexual cycle for their transmission. Chief among these are the mitochondria. Mitochondrial genomes are independent of and quite distinct from the nuclear genome. Mutations in mitochondrial genes result in particular phenotypes in both *Neurospora crassa* and *Saccharomyces cerevisiae*, which arise through loss of mitochondrial function. The mitochondrial genome contains genes for mitochondrial ribosomal RNAs and at least some of the proteins of the

respiratory chain. In other words, the organelle genome specifies some of the organelle proteins, but not all of them. The rest of the organelle proteins are coded by nuclear genes, synthesized in the cytoplasm, and then transported into the mitochondrion.

Genetic maps of mitochondrial genomes, and physical maps derived from restriction analysis, are circular. In many cases, however, the organelle will have linear versions of the genome coexisting with the circular one. *S. cerevisiae* can have more than 100 genomes per mitochondrion, corresponding to about 6500 in each cell. When segregation of organelle genes is followed in genetic crosses, oddly enough, the segregation patterns are consistent with there being only one copy of the mitochondrial genome in the cell. The fact that this is clearly not the case indicates that we do not fully understand how organelle genomes are transmitted from parent to offspring.

Vertical mitochondrial transmission (i.e., from one generation to the next) can depend upon whether or not the fungal strains concerned in a cross show differentiation into "males" and "females." In *Neurospora crassa*, phenotypes controlled by mitochondrial genes are generally transmitted through the female, which is the protoperithecial parent. It is not certain whether this is because of the smaller volume of cytoplasm present in a microconidium (as the male). Paternal and even recombinant mitochondria, however, are sometimes found in the sexual progeny. For example, in a cross between two strains of *N. intermedia*, maternal, paternal, and recombinant mitochondrial DNA have been found in the progeny.

Horizontal transmission (between individuals of the same generation) occurs as a result of hyphal fusion. Although complementation and recombination can be detected between mitochondrial genomes, mycelia that contain genetically different mitochondria (called *heteroplasmons* or *heteroplasmic mixtures*) tend to segregate the different mitochondria into different cells. Cytoplasmic segregation is a general feature of organelle genomes (e.g., in the formation of asexual conidia).

Uniparental inheritance of mitochondrial phenotypes has been observed in yeast, which is isogamous (i.e., does not show a male–female differentiation). Again, the mechanism is unknown, but even in the usual biparental inheritance pattern, mitochondrial genomes can segregate in association with mitosis and the consequential budding. End buds frequently contain one parental mitochondrial DNA, whereas later buds are heteroplasmic or contain recombinant mitochondrial DNA.

In filamentous fungi, mitochondria are not thought to be closely associated with the mitotic spindle, so vegetative segregation may simply be a matter of random physical sorting, though nuclear and mitochondrial genes do influence mitochondrial genome transmission, which is also affected by membrane chemistry. In some ascomycetes subcultured for a long time in the laboratory, altered mitochondrial DNAs due to molecular rearrangements have been associated with modifications in mycelial growth in *N. crassa* and *N. intermedia*, and in the cellular growth in yeast.

In *Schizophyllum commune* mitochondria do not migrate with nuclei during dikaryosis. Mitochondrial inheritance has been studied in matings of *S. commune*, *Agaricus bitorquis*, *A. brunnescens*, *Coprinus cinereus*, *Lentinula edodes*, *Pleurotus ostreatus*, phytophathogenic *Armillaria* species, and *Ustilago violacea*. Plasmogamy may result in all these cases in the production of mycelial colonies composed of sectors containing different mitochondrial DNAs (mitochondrial mosaics). In *Agaricus bitorquis*, *A. bisporus*, *Armillaria bulbosa*, *P. ostreatus*, and *U. violacea*, analysis of dikaryons gave evidence for mitochondrial mixing, sometimes followed by recombination between mitochondrial genomes; recombinant mitochondrial DNAs have also been recovered from dikaryons and dikaryotic protoplasts of *C. cinereus*.

Mitochondrial DNA (mtDNA) of yeast usually makes up 18% of the total DNA, but it has a distinctive very high (82%) AT content, so it is relatively easy to separate from chromosomal DNA. The mtDNA is circular, of 25 nm circumference, and comprises about 7.5×10^5 base pairs. In yeast, mtDNA codes for three of seven polypeptides of the cytochrome *c* complex (the rest derive from nuclear genes), four polypeptide components of a mitochondrial ATPase, and one component of cytochrome *b*. Mutations in these genes can produce recognizable respiratory deficiency phenotypes (e.g., *petite* in *Saccharomyces cerevisiae*, *poky* in *Neurospora crassa*) and thereby provide mitochondrial mutants. In addition, although chromosomal genes code mitochondrial ribosomal proteins, mtDNA determines mitochondrial ribosomal RNA (rRNA) and transfer RNAs (tRNAs).

Mitochondrial ribosomes are similar in size to prokaryotic ribosomes and share some other prokaryotic properties. In particular, protein synthesis on mitochondrial ribosomes is inhibited by chloramphenicol, erythromycin, and several other antibacterials that have no effect on cytoplasmic (eukaryotic) ribosomes. As a result, another kind of mutant phenotype due to mitochondrial mutation is resistance to inhibition by mitochondria-specific drugs. Mitochondrial gene sequences are also similar to equivalent genes in prokaryotes. These features encouraged the endosymbiont theory of mitochondrial origin, which envisages mitochondria to be relics of ancient bacterialike organisms that formed a symbiotic association that resulted in the ancestral "eukaryotic" cell.

The basic procedure for making mitochondrial crosses involves making heterokaryons (diploids in yeast) between haploids carrying the mitochondrial markers. Heterokaryosis may be forced with complementary nutritional (nuclear-gene) mutations. After some vegetative growth of the heterokaryon or diploid, diploid daughter cells (of yeast) or spores or hyphal fragments (of filamentous fungi) are plated out, and the resulting colonies are scored for the mitochondrial markers present in the original haploids. Parental and recombinant clones are usually found, and in relative frequencies that depend very much on the nature of the haploid parents. It is also possible to use the heterokaryons for complementation

analysis. If the constituent haploids carry mutations in different mitochondrial genes they are expected to complement each other in heterokaryons formed between them.

Because the mtDNA is circular, viable recombinants can only result from even numbers of crossovers between DNA molecules. The highest recombination frequency between distant markers is about 25%, and (again, because of the circularity) most pairs of markers tend to show about the same frequency. Nevertheless, multiple-point crosses can effectively define the order of the markers round the circle. Of course, although this sort of recombination *in vivo* may remain important to the organism in nature, it has been supplanted by direct physical analysis of the DNA molecule in laboratory studies.

An increasingly important aspect of mitochondrial transmission is their content of plasmids. Isolates of *Neurospora* from nature commonly contain both linear and circular mitochondrial plasmids. Most are cryptic (i.e., neutral) passengers, but some linear plasmids (notably of *Podospora anserina*) insert into mitochondrial DNA and cause mycelial senescence. Most linear plasmids exhibit typical virus characteristics as far as structure, replication, and function are concerned, and even plasmid-free strains may contain plasmid remnants integrated into their mitochondrial DNA. Plasmid DNA sequences generally encode an RNA polymerase and DNA polymerase, or reverse transcriptase that are used to maintain and propagate the plasmid.

Plasmid DNA, however, is responsible for the killer phenomenon in the yeast *Kluyveromyces lactis* by coding for a killer toxin, which kills cells lacking the plasmid (i.e., cells hosting the killer plasmid are immune to the toxin). These plasmids reside in the cytoplasm and have an expression system independent of both nucleus and mitochondrion. *K. lactis* plasmids can be transferred to other yeasts (including *Saccharomyces cerevisiae*), conferring the killer/immunity phenotype. This shows that the plasmids are autonomous replicons, which can be expressed, in a wide range of host yeasts. The *K. lactis* killer plasmid toxin is chemically and functionally different from a killer toxin produced in *S. cerevisiae*, which is encoded by a double-stranded RNA (dsRNA) virus.

Viruslike particles (VLPs) have been observed in many fungi. They are very similar in appearance to small spherical RNA viruses that have been found in filamentous fungi, but there is little evidence that these particles are effective in hypha-to-hypha infection. Many of the observed VLPs are presumably degenerate or defective viruses that can only be transmitted by hyphal fusions. Viruses of *Agaricus bisporus*, however, require different RNA molecules to produce infective particles as though some are defective viruses and others are helper viruses. Eukaryotic cells do not normally replicate RNA, so these functions are presumably demanding for the virus genome so that different classes of virus have evolved, each performing a special, complementary, function.

S. cerevisiae also harbors five retroviruslike elements, Ty1 to Ty5, as transposons able to integrate into the nuclear genome by targeting particular chromatin structures. The first cytoplasmic plasmid to be observed is the so-called 2 μm DNA of *Saccharomyces* (= Ty1). The name refers to the contour length of the circular DNA molecules that has a base composition similar to nuclear DNA and quite different from mtDNA. There can be 50–100 2 μm DNA molecules per diploid cell, amounting to something like 3% of the nuclear DNA. The 2 μm DNA molecules are transmitted in non-Mendelian fashion, independently of both nuclei and mitochondria. The 2 μm circular DNA carries inverted repeat sequences at either end of two different unique sequence segments; this structure implies that it inserts itself as a whole into the yeast chromosome.

Our discussion so far has dealt with nucleic acid molecules that encode features that segregate in a nonmendelian manner. In the final decade of the twentieth century, however, great attention was given (and continues to be given) to a proteinaceous hereditary element called a *prion protein*. The attention devoted to prions derives from their ability to cause diseases in mammals: scrapie in sheep, bovine spongiform encephalopathy (BSE) in cattle, and in humans, kuru and, most importantly, new variant Creutzfeldt-Jakob disease (nvCJD). In these cases, the pathogenic agent is a variant of a normal membrane protein (the prion protein) that is encoded in the mammalian genome. The variant prion protein folds abnormally and causes normal prion proteins to fold abnormally so that the proteins aggregate in the central nervous system and cause the encephalopathy. Fungi also have prion proteins.

We have already referred briefly (Section 2.3) to the infectious *het-s* phenotype in *Podospora anserina*, which may be a prion protein. An even better candidate is the *PSI*$^+$ form of the Sup35p protein in *Saccharomyces cerevisiae*. Sup35p is an essential yeast protein involved in the termination of translation. In the *PSI*$^+$ state, Sup35p adopts a structural conformation that causes it to direct the refolding of native molecules into a form that can aggregate into filaments of discarded nonfunctional protein. This depletes the cytoplasm of functional translation terminator and results in translation errors. This is the *PSI*$^+$ phenotype, which is inherited by daughter cells following budding and is infectious following cell fusion, in which case it propagates by autocatalytic conversion of the normal form of the protein. The part of the Sup35p protein that makes it a prion (the prion determining domain) is a glutamine/asparagine-rich amino-terminal region that contains several oligopeptide repeats. Removal of these repeats eliminates the ability of Sup35p to propagate *PSI*$^+$; expanding the repeat region increases the spontaneous occurrence of *PSI*$^+$.

Although deleting the analogous repeats from BSE prion protein does not prevent prion propagation and transmission in experimental mice, expansion of the repeat region does increase the spontaneous appearance of spongiform encephalopathies by several orders of magnitude in humans.

It seems very likely that it is the oligopeptide repeats that gives the prion protein the intrinsic tendency to acquire a conformation that enables the protein to refold and effectively polymerize with sister molecules. Database searches for regions with amino acid content comparable to the yeast prions has revealed numerous such domains in eukaryotes, but these are, significantly, lacking from prokaryotes. We urgently need to know whether these other eukaryotic proteins can also behave as elements of protein-based inheritance like prions. Several other human diseases are associated with (and might be caused by) protein misfolding: Huntingdon's Chorea, ataxias, Parkinson's disease, Alzheimer's disease. The proteins involved also have repeating motifs. Further analysis of fungal prions may help in understanding such diseases, but beyond this, there is the question of prion biology. Ongoing work shows that the yeast Sup35p prion provides a selective advantage under adverse conditions, possibly by producing phenotypic variants. These arise because the SUP35 gene encodes translation release factor 3, but PSI^+ aggregation results in inefficient termination of stop codons in the PSI^+ state. This might imply that the prion state is retained because it aids evolution.

Publications and Websites Worth a Visit

Bos, C.J. (1996). Somatic recombination. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 73–95. Marcel Dekker, Inc.: New York.

Davis, R.H. (2000). *Neurospora: Contributions of a Model Organism*. Oxford University Press: Oxford, U.K.

Fungal Genetics Stock Center website at http://www.fgsc.net/methods.

Harris, S.D. (2001). Genetic analysis of ascomycete fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 47–58. Oxford University Press: Oxford, U.K.

Martinelli, S.D. & Kinghorn, J.R. (1994). *Aspergillus: 50 Years on*. Elsevier: Amsterdam.

Osiewacz, H.D. (1996). Extrachromosomal and transposable genetic elements. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 177–208. Marcel Dekker, Inc.: New York.

Osiewacz, H.D. (1996). Genetic analysis of senescence in *Podospora anserina*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 317–335. Marcel Dekker, Inc.: New York.

Pál, C. (2001). Yeast prions and evolvability. *Trends in Genetics* **17**, 167–169.

Perkins, D.D., Radford, A. & Sachs, M.S. (2000). *The Neurospora Compendium: Chromosomal Loci.* Academic Press: Orlando, FL.

Silar, P. & Daboussi, M.-J. (1999). Non-conventional infectious elements in filamentous fungi. *Trends in Genetics* **15**, 141–145.

Strathern, J.N., Jones, E.W. & Broach, J.R. (1981). *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance*. Cold Spring Harbor Laboratory Press: New York.

Swart, K. (1996). Mapping and breeding strategies. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 235–246. Marcel Dekker, Inc.: New York.

True, H.L. & Lindquist, S.L. (2000). A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483.

Tuite, M.F. & Lindquist, S.L. (1996). Maintenance and inheritance of yeast prions. *Trends in Genetics* **12**, 467–471.

Walser, P.J., Hollenstein, M., Klaus, M.J. & Kües, U. (2001). Genetic analysis of basidiomycete fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 59–90. Oxford University Press: Oxford, U.K.

Wolf, K. (1996). Mitochondrial genetics of *Saccharomyces cerevisiae*: the way from genetic crosses to transposable elements. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 247–257. Marcel Dekker, Inc.: New York.

Historical Publications Worth Knowing About

Clutterbuck, A.J. (1974). *Aspergillus*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 447–510. Plenum Press: New York.

Esser, K. (1974). *Podospora anserina*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 531–551. Plenum Press: New York.

Gans, M. & Prud'homme, N. (1976). Chromosome maps of *Coprinus radiatus*. In *Handbook of Biochemistry*, third edition, *Section B*, *Nucleic Acids*, vol. 2 (G.D. Fasman, ed.), pp. 762–764. CRC Press: Cleveland.

Moore, D. (1967). Four new linkage groups in *Coprinus lagopus*. *Genetical Research* **9**, 331–342.

Mortimer, R.K. & Hawthorn, D.C. (1966). Genetic mapping in *Saccharomyces*. *Genetics* **53**, 165–173.

Radford, A. (1976). Chromosome maps of *Neurospora crassa*. In *Handbook of Biochemistry*, third edition, *Section B*, *Nucleic Acids*, vol. 2 (G.D. Fasman, ed.), pp. 739–761. CRC Press: Cleveland.

Sherman, F. & Lawrence, C.W. (1974). *Saccharomyces*. In *Handbook of Genetics*, vol. 1 (R.C. King, ed.), pp. 359–393. Plenum Press: New York.

CHAPTER 6

Mechanisms of Recombination

Revision Concepts

- Examination of a large number of asci reveals a minority of aberrant segregations that show uncharacteristic ratios of ascospores due to localized conversions.
- Conversion is the transfer of a particular genetic sequence from one chromatid to a homologous chromatid during meiosis.
- Flanking markers are loci on either side of and linked to the gene of interest that can reveal any correlation between reciprocal recombination and gene conversion.
- Crossing over tends to favor one end of the conversion segment.
- The theory of eukaryotic (or general or homologous) recombination is based on the fact that 5:3 segregations in octads can be explained by the mechanics of formation of heteroduplex-DNA.
- Homologous recombination is responsible for meiotic crossing over and for general recombination in prokaryotes and eukaryotes.
- Proteins involved in homologous recombination have been widely conserved.
- Enzymes that organize homologous recombination show no sequence specificity, but enzymes involved in site-specific recombination do function via particular DNA sequences.
- Both types of recombination involve similar physical interactions between the DNA molecules.
- Homologous recombination can be used in genetic manipulations involving insertion of recombinant molecules and in gene disruption experiments (reverse genetics).
- Site-specific recombination can be used for chromosomal rearrangements in bacteria, yeast, and all other eukaryotes, transposon transposition, and integration of bacteriophage into host bacterial chromosomes.
- Key steps in homologous recombination are exchange of DNA strands, migration of the crossover point, and resolution of the crossed DNA strands.

- The Holliday model is the most satisfactory mechanistic basis for our understanding of homologous recombination.
- The central feature of the Holliday model is the formation of hybrid- or heteroduplex-DNA regions in which single strands of opposite parental origin are brought together.
- Resolution of the Holliday junction requires two further breaks in the DNA: if the further breaks occur in the two strands that have not so far been broken, a full crossover results; if the same two strands are broken again, no reciprocal crossing over occurs, but the helices exchange regions of heteroduplex-DNA.
- The aberrant 4:4 segregation ratio is direct evidence for the formation of heteroduplex-DNA and is the basis from which all the other aberrant ascus octad ratios can be derived.
- The most satisfactory modification of the Holliday model is the Meselson-Radding model, which suggests that the first single strand formed invades its (still intact) homologous helix.
- No eukaryotic enzymes involved in homologous recombination have yet been characterized.
- Bacterial enzymes involved in homologous recombination are nucleases (*rec* gene products), single-stranded DNA binding proteins (SSB), and Ruv proteins that interact with Holliday junctions produced by the Rec proteins.
- Genetic exchange in yeast involves another double-stranded breakage model.
- Double-strand breaks, chromatid pairing (including synaptonemal complex formation), and formation of recombinants occur in that order during meiosis.
- Interallelic recombination involves production of wild-type progeny from recombination between two mutants that contain complementary point mutations within a gene.
- Genetic recombination provides genotype variability but also the means to eliminate an unfavorable allele without adverse effect on other genes.
- Chromosomes provide a positive selection advantage and a solution to the logistical nightmare of replication and distribution during cell division.
- Crossovers serve a mechanical function during meiosis by stabilizing homologous chromosome pairs during the division process.

6.1 Gene Conversion

The importance of the octad and tetrad segregations that we discussed in the last chapter extends far beyond their contribution to the mapping of chromosomes. Their study reveals phenomena that depend on the molecular details of the process of recombination. Tetrad analysis enabled us to

start understanding recombination, and only fungal octads were able to do this. Examination of large numbers of asci segregating a single gene difference (e.g., black/white spores) will certainly reveal that the majority fall into the categories we illustrated in Fig. 5.3, but there will always be a minority that, instead of showing the usual 4:4 segregation of the two alleles of the gene, have ratios of either 6:2 (e.g., 6 black spores to 2 white spores in the same octad) or 2:6 (i.e., 2 black spores to 6 white spores). These are called *aberrant segregations* because they seem to imply deviation from the Mendelian expectation that alleles segregate in equal numbers at meiosis.

This aberration is called *gene conversion* because a 6:2 ratio might be derived from a 4:4 ratio if one allele is "converted" to an extra copy of the other as the daughter chromosomes emerge from meiosis and before the postmeiotic mitosis. Experiments show that the extra mutant alleles that appear in 6:2 segregations do not arise by some sort of random mutation because they are exactly the same as the original mutant alleles. That means they don't simply have the same phenotype; they have exactly the same DNA sequence. Similarly, where the segregation is 6 wild types to 2 mutant, the extra wild-type alleles are always true wild type and not just another kind of mutant.

Another point is that conversion is quite localized. When several mutant sites are included in the analysis, those that map close together in the same gene usually undergo conversion at the same time (co-conversion). Widely separated mutant sites within the same gene have a low frequency of co-conversion; they usually undergo conversion independently of each other. Sites in neighboring genes are rarely co-converted, which shows that gene boundaries are less important than the size of the conversion segment.

Thus, the picture we have of conversion so far is that it is a transfer of a particular genetic sequence from one chromatid of the four present during early meiosis to a homologous chromatid. The conversion segment replaces the resident sequence and may be a substantial part of a gene, but it is usually less than a whole gene, probably consisting of hundreds rather than thousands of DNA nucleotides.

Conversion is obviously a fairly drastic process, but it occurs with reasonable frequency. The frequency varies greatly from one species to another; several percent of asci of *Saccharomyces cerevisiae* and *Ascobolus immersus* may be conversion asci, whereas the frequency may be less (sometimes much less) than 1% in *Neurospora crassa* and *Sordaria fimicola*. The availability of a good collection of mutations affecting ascospore color in *A. immersus* and *Sordaria fimicola* and *Sordaria brevicollis* made an enormous contribution to the study of gene conversion because ascus segregations from spore-color mutant × wild-type crosses can be observed so easily with a low-power microscope. These studies soon revealed more categories of aberrant segregations in which only half a chromatid (i.e., one spore of a pair in the eight-spored octad) is converted, giving 5:3 or 3:5 ratios. These can occur with the same frequency as the 6:2 (or "whole-

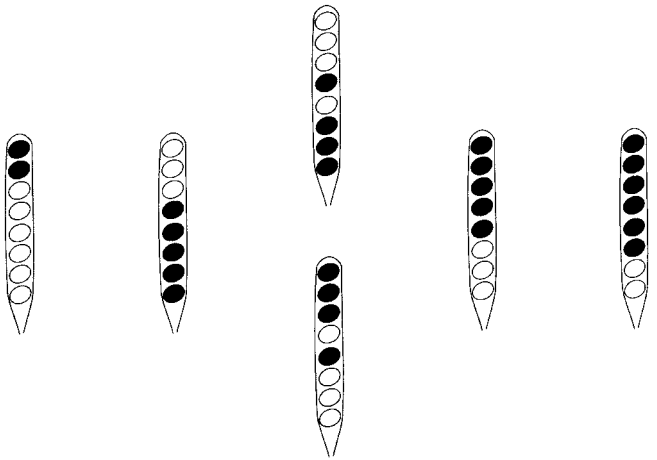


Fig. 6.1. The various sorts of spore patterns in "conversion asci" (ascus octads showing aberrant segregations) derived from spore-color mutant × wild-type crosses that first directed thinking toward the molecular mechanism of recombination.

chromatid") conversions. A 4:4 ratio was sometimes obtained in which two of the four spore pairs are mismatched. This is an aberrant 4:4 ratio, or a postmeiotic segregation 4:4 ratio. The various sorts of aberrant segregations, which are commonly observed in asci derived from spore-color mutant × wild-type crosses, are illustrated in Fig. 6.1.

6.2 Hybrid- or Heteroduplex-DNA

These conversion asci were of enormous importance in the development of theories to explain recombination, but our explanation so far has given no clear reason to associate conversion with recombination, so bear with us for a moment. We have a series of observations from which we can make a series of deductions, so let's see where those deductions lead us.

The 6:2 aberrant segregations allow us to deduce that in maybe 1% of meioses a conversion process replaces a DNA sequence in one chromatid with the sequence present in its homologous chromatid. The 5:3 aberrant segregations imply that the chromatids that emerge from meiosis are structures that can be divided into two. This means that conversion can affect half a chromatid, so the chromatid can be divided into two halves. Remember that the eight spores of an ascus result from a mitotic division immediately following the meiotic division. Mitosis in each of the four haploid meiotic daughter nuclei divides the four chromatids that emerge from the meiosis into the eight daughter chromatids segregated between the eight spores of the ascus.

You can only get a segregation of 5:3 if one of the mitotic divisions segregates one wild-type spore and one mutant spore. Mitosis, however, is supposed to duplicate, not to segregate. The mitotic segregation can only happen if the chromatid that emerges from the meiosis already carries both wild-type and mutant sequences. The simplest assumption to make is that the chromatid consists of a single DNA molecule; then a "half-chromatid" corresponds to one of the DNA strands in the double helix. You would normally expect one "half-chromatid" to be a complementary copy of its partner "half-chromatid" (and therefore to carry the same genetic information). If gene conversion replaces DNA, however, then it is feasible to deduce that gene conversion can sometimes (maybe if it's not completed in time) replace only one strand of a DNA molecule rather than both strands.

If that happens, the chromatid that emerges from meiosis would contain heteroduplex-DNA (also called *hybrid-DNA*), which consists of one strand of mutant sequence replaced into a double helix with one strand of wild-type sequence. The mutant may differ from wild type at only one base pair position; the rest of the few hundreds of DNA nucleotides in the conversion segment from the mutant will be fully complementary to the wild-type sequence, so any problems with base pairing are going to be minor ones. When this hybrid-DNA enters the mitotic division, a perfectly normal mitosis replicates the hybrid-DNA molecule to produce one daughter DNA molecule with fully wild-type sequence and one with fully mutant sequence. This line of argument is beginning to give us a feel for the molecular events involved in conversion, but the relationship with recombination emerged from crosses that included genetic markers placed either side of the alleles involved in gene conversion.

Inclusion of other genes in these spore-color mutant × wild-type crosses enables joint segregation and recombination between the genes to be studied. The ideal arrangement uses two other genes that are linked to, and on either side of, the spore color mutant. These are called *flanking markers*. The flanking markers always show the reciprocal recombination we expect to occur between linked genes, and there is a definite correlation between recombination of flanking markers and gene conversion. For one thing, gene conversions are accompanied by a much higher incidence of crossing over of the flanking markers than would be expected if the two events (i.e., conversion and recombination) were separate and independent of each other.

For example, suppose other data show a map distance of 10% recombination for the markers flanking the spore color gene. This would mean that 20% of asci would show recombination between the genes (see Section 5.5). If conversion and recombination were separate events, only 20% of randomly chosen conversion asci would be expected to show crossing over between the flanking markers. In practice, about half the conversion asci show such a crossover, which is a very strong positive correlation. Further,

the converted chromatid is always one of the two involved in the crossover, and there is a positional relationship between the conversion event and the crossover. This latter feature can only be detected when two alleles of a gene are crossed together, which is a very demanding type of analysis. When such an experiment is done, however, conversion of the left-hand allele is accompanied by crossing over between that allele and the left-hand flanking marker, and conversion of the right-hand allele is accompanied by crossing over between the right-hand allele and the right-hand flanking marker. Such observations suggest that crossing over tends to occur at one end of the conversion segment.

6.3 A Basic Mechanism for Recombination

The notion that 5:3 segregations could be explained by heteroduplex-DNA was the basis for the theory explaining eukaryote recombination. The process has come to be called *general* or *homologous recombination* because it depends on a substantial degree of homology (the longer the better) between the interacting DNA molecules. The enzymes that organize homologous recombination show no sequence specificity. Another form of recombination involves DNA molecules that have only short homologous regions and relies on enzymes specific to particular DNA sequences; this is called *site-specific recombination*. Homologous recombination is the most important recombination process because it is responsible both for meiotic crossing over as well as for general recombination in prokaryotes and eukaryotes. Homologous recombination can also be used in genetic manipulation to target particular regions of the chromosome for insertion of recombinant molecules; for example, in gene disruption experiments (so-called reverse genetics).

Site-specific recombination is used for chromosomal rearrangements in bacteria, yeast, and humans, transposon transposition, and integration of bacteriophage into host bacterial chromosomes. Although they differ in terms of site specificity, both types of recombination involve similar physical interactions between the molecules involved. For homologous recombination there are three key steps: exchange of DNA strands between homologous molecules, migration of the crossover point (also known as the Holliday junction, see later), and resolution of the crossed strands to re-establish intact and physically independent DNA helixes. Although incomplete, current evidence about the proteins involved indicates that the mechanism of homologous recombination has been conserved from bacteria to mammals.

Robin Holliday suggested the model that provides the most satisfactory mechanistic basis for our understanding of homologous recombination, in the mid-1960s. The basic Holliday model views the four-strand stage of meiosis as an association between four DNA double helixes. Two of these,

homologous but nonsister, helices interact in the recombination event. The first step is the occurrence of single-strand breaks (i.e., broken phosphodiester bonds or "nicks") at approximately corresponding points in molecules of the same polarity in both of the nonsister homologues. The breaks allow the free ends to move, and each strand leaves its partner and crosses over to pair through base pairing with the complementary (unbroken) strand of the homologous chromatid. This reciprocal exchange of single strands forms a crossed configuration. It has been given a number of names: because it is a crossover involving single strands it has been called a *half-chiasma*; the connected pair of DNA molecules is also known as a *joint molecule*, and the point where a strand of DNA crosses from one duplex to the other is called the *recombinant joint* or *χ-structure* (Fig. 6.2). There is no doubt that these crossed-strand structures occur during recombination between DNA molecules. Although difficult to see in the long and complex eukaryotic chromosomes, they have been observed in electron micrographs of recombining bacterial plasmids and viruses.

The region where each duplex consists of one strand from each of the nonsister homologous chromatids is the hybrid- or heteroduplex-DNA. Like most aspects of DNA molecules, the recombinant joint structure is dynamic. The branching point is able to move (it's called *branch migration*) along the duplexes. All that is required is that the two DNA helices rotate, and then the strand that unwinds from one molecule is wound up onto the other. Branch migration theoretically could move the point of crossing over for any distance in either direction by rotation of the double helices about their long axes in the same direction, a process that could occur spontaneously by thermal diffusion. *In vivo*, however, enzymes must be involved for any extensive branch migration.

The branching point actually connects two nonsister chromatids together, so the joint molecule formed by strand exchange must be separated, or resolved, before the homologous chromosomes can be disjoined properly at the first division of meiosis. Resolution requires breaks in a further two phosphodiester bonds, with one in each of two DNA strands, but the outcome depends on which two strands are nicked. If the resolving nicks are made in the two strands that have so far remained intact throughout (i.e., in the two molecules that did not begin the strand exchange), then the half-chiasma is converted into a full-chiasma with reciprocal-strand exchange in a full crossover that contains a stretch of heteroduplex-DNA. This crossover is a conventional recombination event for any flanking markers, and any mutant alleles within the heteroduplex region are candidates for gene conversion (Fig. 6.3).

The alternative way of resolving the half-chiasma is by nicking the same two strands involved in the original strand exchange. This releases the original parental duplexes, which remain intact (i.e., without a reciprocal crossover) except that a short segment of DNA is transferred between the two molecules. Each parental duplex is consequently left with a length of

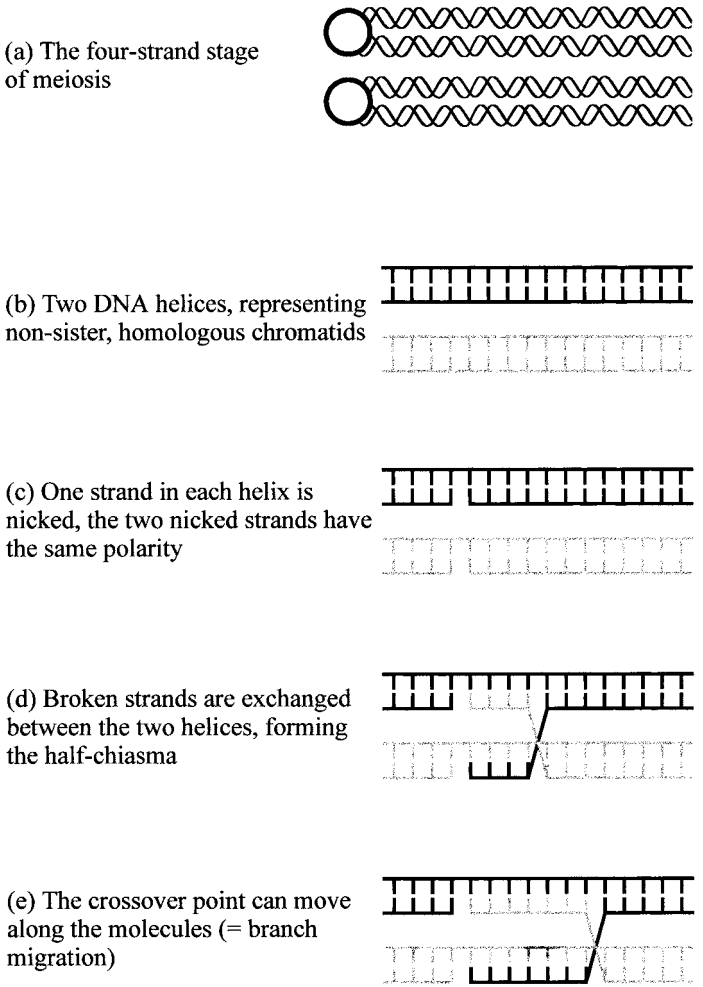


Fig. 6.2. Diagrams of the Holliday model for the initiation and migration steps in the homologous recombination process. (a) is a reminder that although all the subsequent diagrams are greatly simplified, genetic recombination is a reciprocal exchange between nonsister chromatids at the four-strand stage of meiosis, and each chromatid consists of a DNA double helix. In this diagram, the large circles represent centromeres. (b) shows the starting point for our diagrams; namely, linearized DNA helixes of the two chromatids involved in crossing over (remember that there are two other chromatids, which we are ignoring because they are not involved in this crossover). In (d) the Holliday junction structure has been formed. In this configuration single strands form a crossed configuration. It has been given a number of other names since Robin Holliday first postulated its occurrence. It is a crossover involving single strands and has been called a *half-chiasma*; the connected pair of DNA molecules is also known as a *joint molecule*, and the point where a strand of DNA crosses from one duplex to the other is called the *recombinant joint* or χ *form* (and see Fig. 6.3). (e) shows that the crossed-strand structure can migrate along the DNA molecules; here migration for only a few bases is shown, but in reality the structure can migrate many hundreds or thousands of bases. (Based on Fig. 14.2 in Lewin (2000), *Genes VII*, Seventh edition, Oxford University Press.)

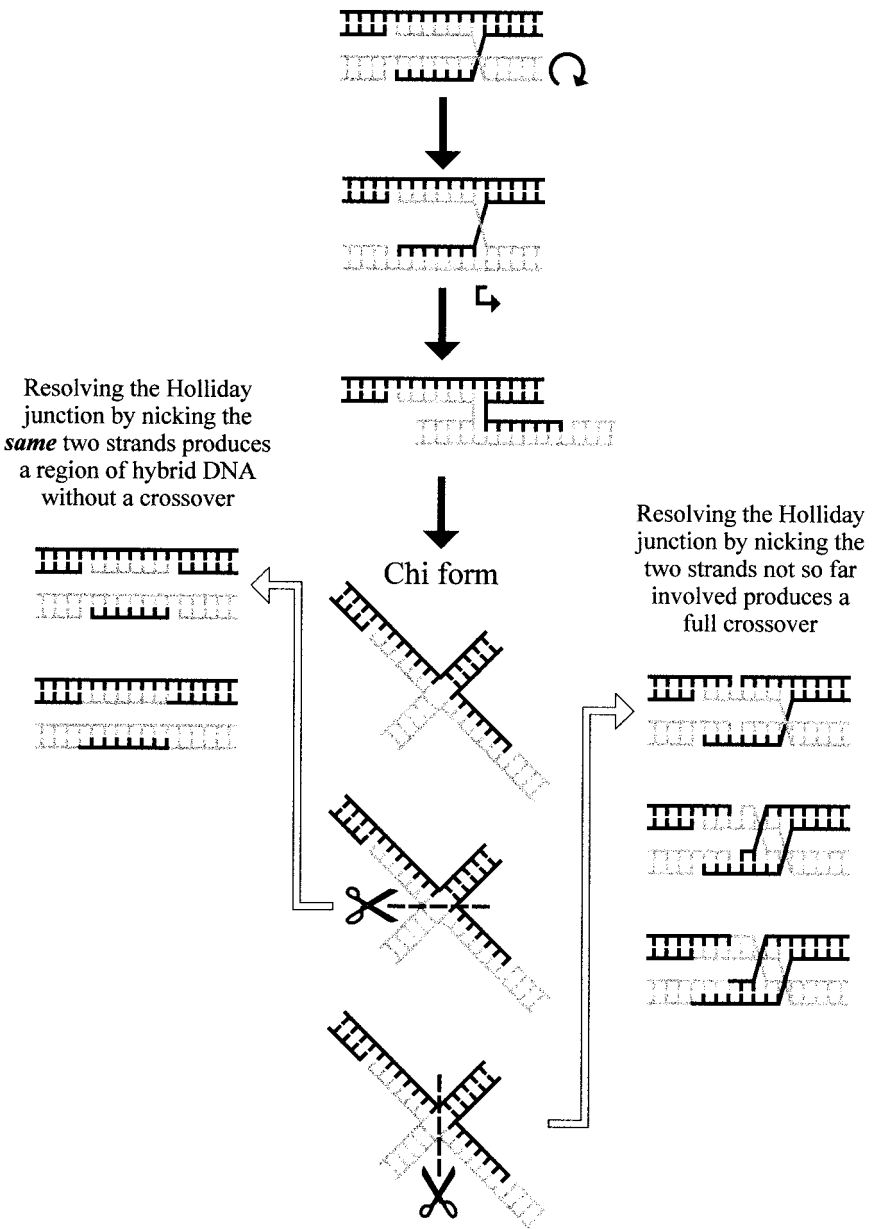


Fig. 6.3. The central set of diagrams explain how the χ form is derived from the more conventional diagrams by rotating one chromatid around its long axis and then around the crossover point. The χ form is important in the enzymatic resolution of the Holliday structure. If further nicks are made in the two strands that have so far remained intact throughout, the half-chiasma is converted into a full-chiasma with reciprocal strand exchange in a full crossover, which contains a stretch of heteroduplex-DNA (bottom χ form diagram and diagrams on right). On the other hand, further nicks in the two strands that have already been broken produce a region of hybrid DNA without a crossover (penultimate χ form diagram and diagrams on left). (Based on Fig. 14.4 in Lewin (2000), *Genes VII*, Seventh edition, Oxford University Press.)

heteroduplex-DNA equal in length to the distance migrated by the branching structure and which is a candidate for gene conversion if it includes a mutation distinguishing the two duplexes (Fig. 6.3). All of the broken phosphodiester bonds can subsequently be repaired by a DNA ligase activity to restore the structural integrity of the chromatids, which will then be able to disjoin properly in the meiotic divisions.

The Holliday model for homologous recombination successfully accounts for recombination between double-stranded DNA molecules. In at least a proportion of cases, it achieves the breakage and reunion with physical exchange of parts between different but homologous DNA molecules essential for reciprocal crossing over without threatening fragmentation of the chromatid. Although both strands of both DNA helices must be broken to accomplish a reciprocal crossover, the Holliday model achieves this by having breaks in single strands that occur in succession. A central feature of the model is formation of heteroduplex-DNA regions in which single strands of opposite parental origin are brought together. If the heteroduplex includes a mutant site that was heterozygous in the original cross, then there will be a base mismatch within the heteroduplex-DNA. For example, if the wild-type base pair is GC and this is changed to AT in the mutant, then there will be an AC mismatch in the heteroduplex-DNA on one chromatid and a GT mismatch in the heteroduplex-DNA on the homologous chromatid.

Remember that these events occur at the four-strand stage of meiosis, before the first division of meiosis. If nothing else happens other than the completion of meiosis and the formation of the ascus, therefore, then a postmeiotic segregation will happen to two of the four meiotic products, giving the aberrant postmeiotic 4:4 segregation ratio (Fig. 6.4). This aberrant 4:4 segregation ratio is direct evidence for the formation of heteroduplex-DNA. It is the basic aberrant segregation ratio from which the other aberrant ratios observed in ascus octads can be derived by correction of mismatches in either wild-type to mutant or mutant to wild-type directions (Fig. 6.4).

6.4 Correction of Base Mismatches in Heteroduplex-DNA Generates Aberrant Segregation Ratios

Let us return to our example to see how correction of base mismatches in heteroduplex-DNA can explain the other aberrant octad ratios. In our example the wild-type base pair is GC and the mutant is AT, so there will be an AC mismatch in one heteroduplex and a GT mismatch in the other. Both of the mismatches can be recognized as such by repair enzymes, and both could be corrected to either AT or GC. Remember that these mismatched heteroduplexes are now located in physically different chromatids, so there is no reason to expect that what happens to one will necessarily

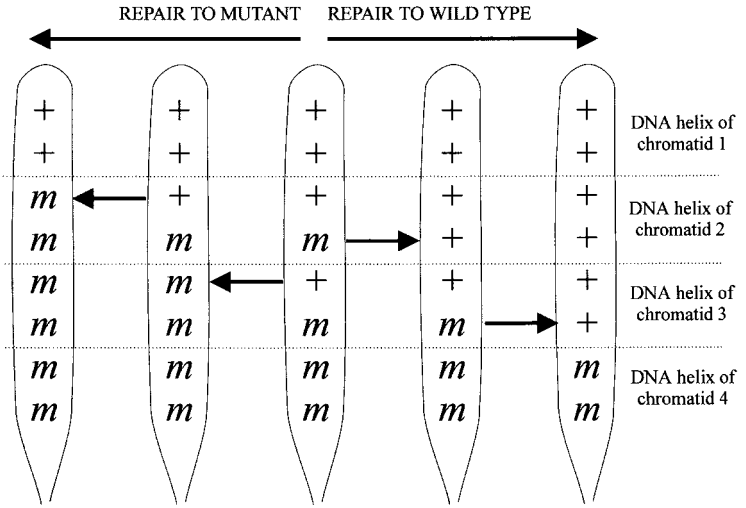


Fig. 6.4. Potential for mismatch repair to generate 5:3 and 6:2 octad ratios from the aberrant 4:4 segregation ratio, the basic original pattern shown in the central ascus.

happen to the other. The possibilities are illustrated in Fig. 6.4. Correction of both chromatids to wild type (i.e., to GC) will give six wild-type to two mutant spores; correction to mutant (i.e., to AT) on both chromatids, two wild-type to six mutant spores. If one chromatid corrects to wild type and the other corrects to mutant, an apparently normal 4:4 ratio will emerge (i.e., the octad segregation will not give any hint of the events that produced it). Correction of one chromatid only will give 5:3 or 3:5 ratios, depending on the direction of the correction. Finally, as stated earlier, when correction does not occur at all, the result is the aberrant 4:4 postmeiotic segregation ratio (Fig. 6.4).

A number of repair systems recognize mispaired bases in DNA, and replace a segment of one of the strands to restore complementary pairing. Substrates for the repair systems include mispaired bases in heteroduplex-DNA formed during recombination, as well as mutations and replication errors. Radiation-sensitive mutants in yeast have been used to deduce the genes involved in DNA repair functions. As a result, they are called *RAD genes* and they fall into three general groups. The RAD3 group is involved in excision repair of mutagen damage, the RAD6 group is required for repair of replication-induced damage, and the RAD52 group is concerned with recombination repair. Enzymes that scrutinize the DNA, searching for apposed bases that are incorrectly paired, accomplish repair of mismatches caused by recombination. The correction process starts with a single-strand break made close to the site of the mismatch. A nuclease then digests away a segment of the nicked strand across the site of the mismatch and a short

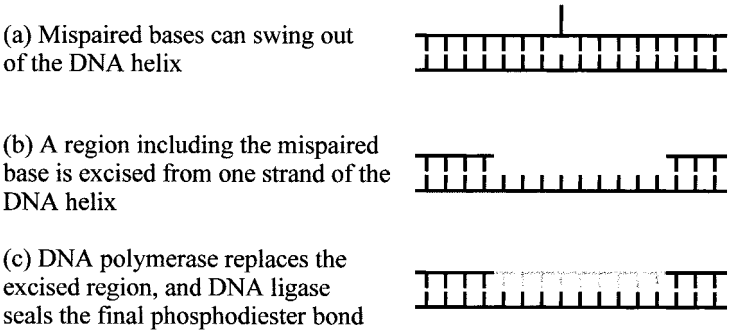


Fig. 6.5. Repair of mismatches caused by recombination is done by enzymes that scrutinize the DNA searching for bases that are incorrectly paired. The nuclease digests away a segment of the strand across the site of the mismatch and extending beyond it. Finally, the gap created by this excision is filled in by repair synthesis and the single-strand ends are rejoined.

segment extending beyond it. Finally, the gap created by this excision is filled in by repair synthesis and the single-strand ends are rejoined (Fig. 6.5).

6.5 Modifying the Basic Recombination Model

The Holliday model is, at least theoretically, clearly, capable of accounting for all of the observed features. It encounters difficulties, though, when predictions about relative frequencies of the various outcomes are compared with observation. The basic model as described earlier claims that heteroduplex-DNA is formed on both chromatids involved in the events (i.e., heteroduplex-DNA is formed symmetrically). This would lead to the prediction that both chromatids should be equally able to show conversion (in either direction) or no conversion.

Observations made of yeast, *Sordaria*, and *Ascobolus*, however, reveal a strong tendency for one chromatid to retain its original configuration, whereas the other undergoes conversion or becomes internally heteroduplex. For example, a very low frequency of aberrant postmeiotic 4:4 segregation ratios are observed in yeast, and reciprocal conversion (one mismatch corrected to wild type; the other to mutant) is very rare in yeast except when flanking markers are recombined. In *Ascobolus immersus* and yeast, asci with the ratio five wild type to three mutant predominate to the extent that there are virtually no five mutant to three wild-type ratios, and *Sordaria* shows a similar imbalance in 5:3 ratios, although less extremely. These observations can be accommodated if it is presumed that the

formation of heteroduplex-DNA is *not* symmetrical, and that, at least at first, only one chromatid exchanges a strand of its DNA double helix.

The most satisfactory modification is called the *Meselson–Radding model* after its proposers. This suggests that a single-stranded nick occurs in one of the double helices and a resultant free end of a DNA molecule invades the homologous (and unbroken) double helix, displacing one of its strands so that the invading molecule can base pair with the other (Fig. 6.6). The displaced strand is degraded and the single-strand gap in the donor chromatid is repaired by reference to the complementary surviving strand. The outcome is the unilateral (or asymmetric) production of heteroduplex-DNA, which would account for the observed shortages in some of the aberrant segregation ratios.

The Meselson-Radding model also copes with another difficulty of the basic Holliday model, which is that the latter seems to require single-strand nicks to appear at the same time in equivalent positions in two different DNA helixes. The Meselson-Radding modification envisages that only one phosphodiester bond need be broken to initiate the process; all other breaks (and DNA degradation or repair) are consequences of the behavior of the invading DNA molecule and therefore make for easier accounting. It is important, however, to recognize that the Meselson-Radding model is a modification of the Holliday model, not a replacement. Although the Holliday model, which allows for the formation of symmetrical

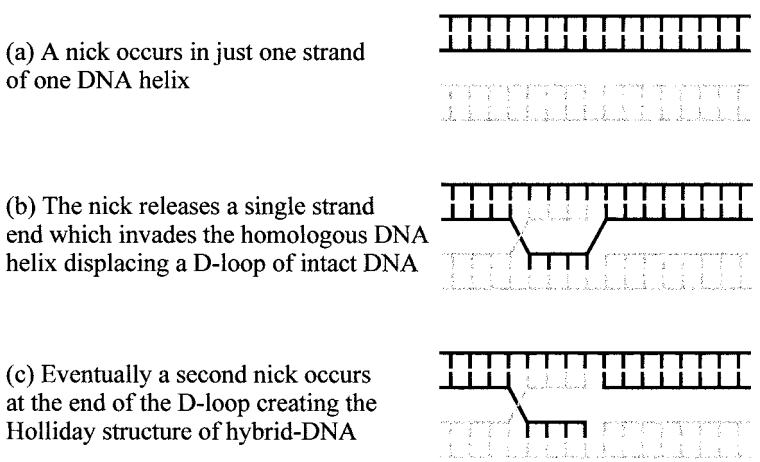


Fig. 6.6. The Meselson-Radding model suggests that a single-stranded nick occurs in one of the double helixes and a resultant free end of a DNA molecule invades the homologous (and unbroken) double helix, displacing one of its strands so that the invading molecule can base pair with the other. This avoids the necessity for postulating adjacent breaks in two separate chromatids. (Based on Fig. 13.25 in Brown (1999), *Genomes*, BIOS Scientific Publishers.)

heteroduplex-DNA exclusively, is ruled out by the inequalities in the relative frequencies of aberrant ratios seen in some species, the other side of the coin is that the frequent occurrence of equal ratios in other fungi, including strains of *Sordaria* and *Ascobolus*, also decisively rules out the asymmetrical process as an exclusive model. The power of the basic Holliday model is that it can be modified, and that it is easy to contemplate the modifications being determined by local differences along chromatids, between chromatids, and/or between strains and species.

The only enzyme activities involved in homologous recombination that have been characterized so far unfortunately occur in the bacterium *Escherichia coli*, coded by the *rec* and *ruv* genes. One of these activities (to which three *rec* genes contribute) is a nuclease that produces a single-strand nick at the sequence 5′-GCTGGTGG-3′ (i.e., the χ sequence). Such sequences occur at about 5 kb intervals in *Escherichia coli* and provide hotspots for recombination. The complex of RecB, C, and D proteins binds to double-stranded DNA and unwinds it to produce single-stranded tails. Another activity is a single-stranded DNA binding protein (SSB protein) that binds to the free single-stranded end released by nicked DNA and mediates its invasion of the intact partner helix. RecA protein, however, generates synapsis between DNA molecules by polymerizing into a helical cage around single-stranded DNA. This complex can interact with double-stranded DNA and when a region of homology extending more than 30–50 nucleotides is found, strand exchange is initiated, the homologous strand being displaced by the 3′-end of the invasive molecule.

Proteins produced by genes of the two *ruv* operons interact with Holliday junctions produced by RecA protein. RuvA, RuvB, and RuvC together form a "recombination machine" that catalyses branch migration and resolution of the Holliday junction into mature recombinant products. RuvA and RuvB together form a molecular motor able to rotate the two DNA helices at the Holliday junction so that the branch point moves. The RuvA protein is a tetramer in which several helix–hairpin–helix DNA binding motifs line the grooves between the monomers. These hold the four arms of the χ form of the Holliday junction. RuvB assembles around the DNA as hexameric rings, which move the DNA along using energy derived from ATP hydrolysis. Branch migration is unidirectional, the direction being determined by the way the RuvB ring is assembled onto the DNA. It seems that RuvB pulls the DNA into and through the RuvA/RuvB complex. When migration is completed, at a 5′-[A/T]TT[G/C]-3′ sequence, the endonuclease RuvC protein (which is a dimer) resolves the Holliday junction structure by cleaving the strand between the T and [G/C] of this recognition sequence.

Although these are bacterial systems, there is some hope that they are representative of homologous recombination in general because the *E. coli* proteins are very similar to the proteins involved in recombination in yeast. There is, for example, a yeast protein called RAD51 that has sequence

similarity with the single-stranded DNA binding RecA protein of *E. coli* and is believed to have the same function. No eukaryotic counterparts of RecB, C, or D, have been found.

Another example of what is presently known about prokaryotic DNA recombination and what might be indicative of the eukaryotic system is the Cre recombinase of bacteriophage P1, which might suggest that the machinery required for homologous recombination is highly conserved. The Cre recombinase from bacteriophage P1 arranges a site-specific recombination reaction to integrate the virus genome to establish the lysogenic state. Structure determination of the Cre/DNA complex shows that the DNA adopts a conformation similar to the four-way Holliday junction, with four Cre molecules, each of which cleaves the DNA to form covalent 3′-phosphotyrosine links with the DNA. Strand exchanges between the Cre molecules first establish a Holliday junction and then resolve this into recombinant products.

At least some, and maybe most, of the genetic exchange in yeast is initiated by breakage across both strands of the double helix. This is a radical process, which is only feasible because the endonuclease enzyme involved forms covalent bonds with the DNA molecules and therefore prevents them from coming apart. Even more radical is that the initial strand breakage is immediately enlarged to a gap by exonuclease digestion of one strand on each side of the breaks, which obviously leads to loss of information from that chromatid. The lost information is retrieved by the 3′-ends this generates by invading the homologous regions in the DNA helix of the homologous (nonsister) chromatid (Fig. 6.7). This invasion displaces one of the resident strands, and the amount displaced is increased as DNA synthesis extends the 3′-end of the invading molecule.

The length of resident DNA displaced is eventually large enough to fill the gap on the original (broken) chromatid and it can anneal with the complementary single-stranded sequences on the far side of the gap. When this stage is reached there is a stretch of heteroduplex-DNA on each side of the gap, and the single-stranded DNA molecule that was displaced from the homologous chromatid represents the gap itself. The gap is repaired by DNA synthesis from the "noninvading" 3′-end, so replacing the information "lost" when the gap was opened up, but leaving the molecule with a Holliday junction crossed-strand structure, or recombinant joint, at each end. These can be resolved in a number of ways: if they are resolved in opposite ways, a reciprocal crossover is produced; if both are resolved in the same way, the chromatids are released without a crossover. In either case heteroduplex-DNA is left at each end of the region involved in the exchange.

The double-strand break model of recombination looks at first sight like a surprisingly complicated modification of the basic Holliday model, particularly because genetic information is destroyed at the start of the process. Errors in retrieving that information could cause massive mutations, and

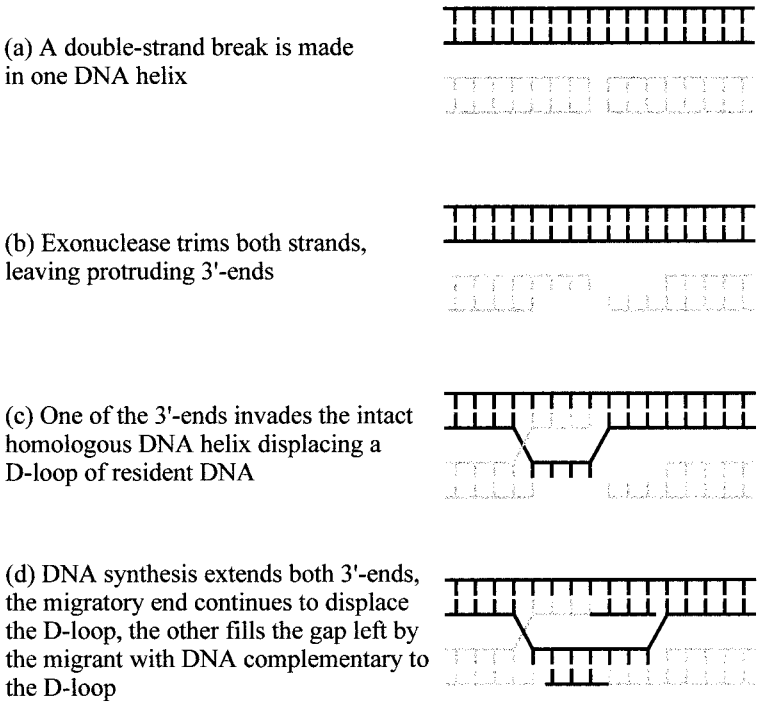


Fig. 6.7. In yeast, genetic exchange is initiated by breakage across both strands of the double helix. The endonuclease enzyme forms covalent bonds with the DNA molecules, which prevents them from coming apart. Initial strand breakage is immediately enlarged to a gap by exonuclease digestion of one strand on each side of the breaks. Lost information is retrieved by extension of the 3′-ends this generates into the homologous regions in the DNA helix of the homologous (nonsister) chromatid. Holliday junction structures are eventually left and can be resolved as shown in Fig. 6.3. (Based on Fig. 13.29 in Brown (1999), *Genomes*, BIOS Scientific Publishers.)

many people are unwilling to accept the notion that DNA undergoes frequent double-strand breaks during meiosis. There is little doubt, however, that it occurs in yeast and may well occur in other organisms, although there is not enough evidence for this. Recombination in eukaryotes, however, is bound to be much more complex than it is in prokaryotes because it is not just naked DNA that interacts; we have to consider the involvement of the whole chromatid, including both DNA and protein. Genetic recombination, synaptonemal complex, and chromatid synapsis are all features of the four-strand stage of meiosis. They are not independent of each other.

Work with yeast mutants that have defects in meiotic events shows that double-strand breaks, chromatid pairing reactions (including synaptonemal complex formation), and formation of recombinants occur in that order. Double-strand breaks appear very early and disappear as synaptonemal

complexes develop. This relative timing suggests that the formation of synaptonemal complexes might be a consequence of the initiation of recombination by the double-strand breaks and their subsequent processing. Recombinants in yeast appear at the end of the pachytene stage of meiosis, so recombination is completed after synaptonemal complexes have been formed. In effect, therefore, it seems likely that double-strand breaks initiate chromosomal synapsis and that the numerous proteins involved in manipulating the DNA during consequential recombination processes contribute to the structure that is observed and called the *synaptonemal complex*. There is not an absolute correspondence between the synaptonemal complex and the recombination machinery, though. On the one hand, mutations that abolish recombination also abolish synaptonemal complex development. Without the initiating event of recombination, the synaptonemal complex presumably cannot form. On the other hand, however, some mutants lacking in synaptonemal complex formation can nevertheless form recombinants. The synaptonemal complex presumably has functions independent of recombination, perhaps in later stages of meiosis.

6.6 Models and the Real World

Two independent point mutations of the same gene (alleles) are likely to be at different sites and able to produce wild-type progeny by recombination when the two mutant strains are intercrossed. This is *interallelic recombination* and the occurrence of recombination between sites within the same gene enables fine structure genetic maps to be constructed that represent the internal structure of the gene. Maps can be made on the basis that recombination frequencies from interallelic crosses are proportional to the distance between the alleles, although the inferred distances are often not very additive.

The best technique to use for mapping closely situated mutation sites is to use small overlapping deletions that cover the region. A point mutation cannot recombine with a deletion to form wild-type progeny if the deletion covers the site of the point mutation. Thus, on the one hand, genetic analysis can be used to define a deletion by the set of point mutants with which it is unable to form recombinants. On the other hand, point mutants that do not recombine with a deletion can be assigned to the segment of gene that is missing in the deletion mutant. Determination of recombination frequencies is not necessary. Rather, it's only a matter of whether or not recombinants are formed. A collection of overlapping deletions can be used to locate a new point mutation anywhere in the region covered by the collection of deletions (Fig. 6.8).

Allele maps can also be established by using flanking markers, on the expectation that recombination between the alleles that produces wild-type recombinants will also recombine the flanking markers; however, the flank-

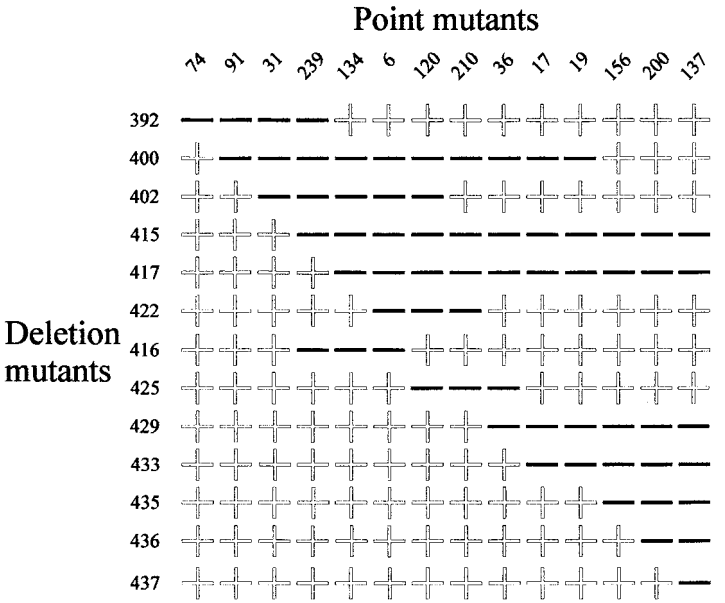


Fig. 6.8. Diagram of deletion mapping in the *cyc*-1 gene of *Saccharomyces cerevisiae*. Each point mutation was crossed with the series of deletion mutants and the appearance of wild-type recombinants was scored: plus = recombinants obtained, minus = no recombinants obtained. The relative lengths and positions of the deletions enable the point mutations to be positioned.

ing marker genotype of intragenic recombinants will depend on the order of the alleles relative to the flanking markers. As a result, new point mutation alleles can be placed in map order by determining the recombinant flanking marker combination most commonly encountered in interallelic crosses. Of course, a prerequisite is a pair (at least) of suitable flanking markers and a set of strains in which the new point mutations have been combined with one or other of the flanking markers (Fig. 6.9).

The flanking marker approach has been widely used and found to give clear left–right relationships of alleles within the gene under study, but sites that are close together give anomalous results. The reason for these, of course, is that interallelic recombinants do not arise from reciprocal crossing over, but from gene conversion. Indeed, polarity in gene conversion can even be used to map alleles within some genes. The first example of this involved some ascospore color mutants of *Ascobolus immersus* in which it was found that when mutant × mutant crosses were made it was always the left-hand site that underwent conversion to wild type (thus producing dark-spored recombinants). An integrated series of crosses enabled the sites to be arranged in map order on this basis.

A map of allele positions established in this way seems to be reliable because, in another case in *Neurospora*, the sequence of mutant sites established by conversion polarity agreed with the order of mutated sites in the amino acid sequences of mutant proteins. Explanation of conversion polarity depends on the events that cause conversion. For example, the heteroduplex-DNA might "spread" into the gene from one side because of unilateral branch migration of the Holliday junction. A gene that is close to a hotspot for initiation of recombination may show strong unilateral migration of heteroduplex-DNA, and consequently strong conversion polarity. A gene located midway between two recombination hotspots, on the other hand, might be subject to bilateral migration of heteroduplex-DNA from the flanking hotspots and, therefore, show no conversion polarity at all. Another aspect of the conversion process, which could also cause conversion polarity, is if the mismatch "repair to wild type" process favors one allele over another on the basis of their relative positions within the gene. This could happen if the repair mechanism is influenced by, or favors, a particular strand of the DNA. Remember that the two strands of a DNA helix run in different chemical directions (one runs 5′ to 3′; the other 3′ to 5′), so this explanation only demands that a nuclease with a particular specificity degrades a strand from the site of mismatch.

Mapping alleles by recombination frequencies has also been successfully accomplished. The only justification for using the frequency of wild-type

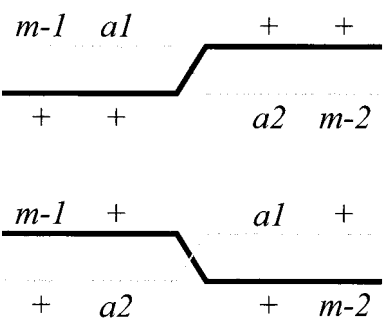


Fig. 6.9. Diagram of allele mapping using flanking markers. The arrangement of the alleles (*a1* and *a2*) is established on the basis of the most-frequently observed flanking markers (*m-1* and *m-2*) observed in progeny recombinant for the wild-type alleles at the *a* locus. We know that the cross is made between an *m-1*, *a1* strain and an *m-2*, *a2* strain. Thus, if the $a1^+$, $a2^+$ recombinants are mostly wild type for *m-1* and *m-2* as well, then the order of the mutant sites must be *m-1*, *a1*, *a2*, *m-2* (top diagram). On the other hand, if the majority of $a1^+$, $a2^+$ recombinants have mutant phenotypes for the flanking markers, *m-1* and *m-2*, then the order of the mutant sites must be *m-1*, *a2*, *a1*, *m-2* (bottom diagram). (Based on Fig. 45 in Fincham, Day & Radford (1979), *Fungal Genetics*, fourth edition, Blackwell Scientific Publications.)

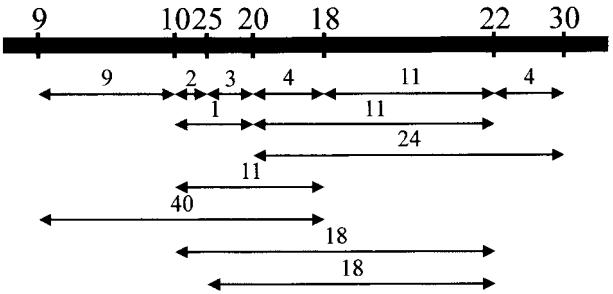


Fig. 6.10. Mapping alleles of the *pan*-2 locus of *Neurospora* by recombination frequencies. Numbers above the line are allele reference numbers. Numbers over the arrows show the prototroph frequency (wild-type progeny per 10^4 viable ascospores) obtained from the pairwise crosses made between the mutants. (Based on Fig. 43 in Fincham, Day & Radford (1979), *Fungal Genetics*, fourth edition, Blackwell Scientific Publications.)

recombinants from mutant × mutant crosses as a mapping criterion is the internal consistency of the maps, and this varies considerably. Lack of internal consistency is no great surprise because the fact is that gene conversion accounts for interallelic recombination and there is no expectation that the frequency of conversion should be a measure of distance between mutant sites; however, the additivity of distances in allele maps based solely on recombination frequencies is often quite convincing (Fig. 6.10). Wild-type recombinants will only occur when one site, but not the other, is converted. Good additivity implies that there is an underlying randomized feature of the conversion process that is related to position and distance between sites. It could be that genes that show good additivity in their allele maps are equidistant between any recombination hotspots, so that heteroduplex-DNA migrates equally from both sides and ends at a random position within the gene. If the probability of heteroduplex-DNA covering one site but not the other in any mutant × mutant cross is randomized, then the frequency of wild-type recombinants produced by conversion could be related to the distance between the sites.

Genetic recombination is generally considered to provide genotype variability. The argument runs that recombining the genes in a chromosome allows new mutants and new gene assortments to be tested against the forces of selection. Favorable and unfavorable mutations can be tried out, and recombination provides the means whereby an unfavorable allele can be eliminated without adverse effect on other genes on the same chromosome. If recombination did not occur, then selection would have to be applied to the entire chromosome.

That, however, is not the whole story. Indeed, if you think through the "provision of genotype variability" story you might well ask, "Why have

chromosomes at all if variability is so important?" If genes were independent entities, each one a minichromosome, for example, then random segregation would generate all the genotype variation the organism could ever need. From viruses on upward, however, organisms do have chromosomes, so there must be a positive selective advantage in having genes on chromosomes. At least part of that advantage is likely to be that chromosomes provide a solution to the mechanical and management nightmare of replicating and distributing the correct number of copies of all the genes to the daughter nuclei at each cell division. An evolutionary balance has been reached between a trend toward stringing genes together in chromosomes to make them easy to distribute and a trend toward breaking the genes apart to provide greater variability.

Even that, however, is still not the whole story. Several features of eukaryote recombination that we take for granted are disturbed in meiotic mutants. Even the distribution of recombination events that are resolved as crossovers is controlled to ensure that the probability of no crossovers occurring between a homologous pair of chromosomes is very low (<0.1%). This serves to remind us that crossovers serve a mechanical function in meiosis. Crossovers stabilize homologous chromosome pairs late into the first meiotic prophase and long after chromosome synapsis has dissipated, providing a counterforce to the contracting spindle fibers and enabling the chromosomes to be separated without tangling. Furthermore, recombination is a "meiotic checkpoint" that is necessary for progress through meiosis. The meiotic division is blocked until recombination has been completed successfully. This is again a matter of ensuring that disjunction is not attempted until recombination has provided the mechanical ties that help to organize chromosome segregation.

Although we almost always visualize recombination as involving two DNA double helices (Figs 6.2–6.4, for example) it is essential to remember that in eukaryotes only about one third of the mass of the organelles involved in recombination, the chromatids, consists of DNA. There is about one-third histone protein and one-third nonhistone proteins. Many of these nonhistone proteins will be the enzymes able to manipulate DNA: changing its conformation, degrading it, synthesizing it—and controlling it.

For example, there is a mutant of yeast that abolishes chiasma interference (one crossover event inhibiting the occurrence of another nearby). This mutation affects synaptonemal complex formation, but it demonstrates categorically that a universal feature of recombination depends on proteinaceous components of the chromosome rather than the DNA. Given the number of proteins involved in recombination, variation is inevitable. There is a high degree of variation in recombination frequency within natural populations. This is often expressed in linkage distances being under the control of genes whose alleles show dominant–recessive relationships. Thus, for example, segregation of alleles of a single gene might distinguish low recombination frequency in a chromosomal region caused by one allele

from high recombination frequency in that interval caused by the other allele.

Our discussion of linkage mapping started with the proposition that if crossing over is randomized along the length of a chromosome, then it will occur more often between two genes that are far apart than it will occur between two genes that are close together. In other words, recombination frequency will be proportional to the distance between genes. There must be some truth in this proposition because linkage mapping works; this is an astonishing fact given the enormous number of proteins involved whose functions could disturb proportionality to distance. There is still a lot to be discovered about recombination!

Publications and Websites Worth a Visit

Brown, T.A. (1999). *Genomes*. BIOS Scientific Publishers: Oxford, U.K.

Craig, N.L. (1988). The mechanism of conservative site-specific recombination. *Annual Reviews of Genetics* **22**, 77–105.

Hiom, K. (2001). Recombination: homologous recombination branches out. *Current Biology* **11**, R278–R280.

Gopaul, D.N. & Van Duyne, G.D. (1999). Structure and mechanism in site-specific recombination. *Current Opinion in Structural Biology* **9**, 14–20.

Guo, F., Gopaul, D.N. & Van Duyne, G.D. (1997). Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* **389**, 40–46.

Haber, J.H. (1998). A locus control region regulates yeast recombination. *Trends in Genetics* **14**, 317–321

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. (1994). Biochemistry of homologous recombination in *Escherichia coli*. *Microbiological Reviews* **58**, 401–465.

Kolodner, R.D. & Marsischky, G.T. (1999). Eukaryotic DNA mismatch repair. *Current Opinion in Genetics & Development* **9**, 89–96.

Lamb, B.C. (1996). Ascomycete genetics: the part played by ascus segregation phenomena in our understanding of the mechanisms of recombination. *Mycological Research* **100**, 1025–1059.

Lewin, B. (2000). *Genes VII*, seventh edition, Oxford University Press: Oxford, U.K.

Mengiste, T. & Paszkowski, J. (1999). Prospects for the precise engineering of plant genomes by homologous recombination. *Biological Chemistry* **380**, 749–758.

Page, A.W. & Orr-Weaver, T.L. (1997). Stopping and starting the meiotic cell cycle. *Current Opinion in Genetics & Development* **7**, 23–31.

Parikh, S.S., Mol, C.D., Hosfield, D.J. & Tainer, J.A. (1999). Envisioning the molecular choreography of DNA base excision repair. *Current Opinion in Structural Biology* **9**, 37–47.

Rice, D.W., Rafferty, J.B., Artymiuk, P.J. & Lloyd, R.G. (1997). Insights into the mechanisms of homologous recombination from the structure of *RuvA*. *Current Opinion in Structural Biology* **7**, 798–803.

West, S.C. (1992). Enzymes and molecular mechanisms of genetic recombination. *Biochemistry* **61**, 603–640.

Historical Publications Worth Knowing About

Case, M.E. & Giles, N.H. (1958). Recombination mechanism at the *pan*-2 locus in *Neurospora*. *Cold Spring Harbor Symposia on Quantitative Biology* **23**, 119–135.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research* **5**, 282–304.

Sigal, N. & Alberts, B. (1972). Genetic recombination: the nature of crossed strand-exchange between two homologous DNA molecules. *Journal of Molecular Biology* **71**, 789–793.

CHAPTER 7

The Physical Genotype

Revision Concepts

- Recombinant DNA technologies detect differences in DNA sequences that result in scorable DNA markers.
- Before DNA markers came into prominence, proteins were used to identify loci and characterize strains.
- Protein or enzyme polymorphisms are called *isozymes* (multiple forms of an enzyme) or *allozymes* (enzyme variants produced by particular alleles).
- Differences in DNA sequence are called *DNA polymorphisms*, which provide the basis for direct physical analysis of the genotype.
- DNA polymorphism advantages over conventional functional mutation are that sequence difference is frequent and detected directly, and no functional phenotype is required.
- DNA polymorphisms that do not affect a phenotype are not subject to selection pressure, so the majority of DNA variation is located in noncoding regions.
- Up to 30% of fungal nuclear DNA is comprised of noncoding sequences.
- DNA sequence differences arise from transitions, transversions, deletions, insertions, and errors in replication.
- Simplest DNA polymorphisms are caused by elimination or creation of a recognition site for a restriction enzyme.
- Restriction fragments are pieces of DNA produced when DNA is digested with a restriction endonuclease that recognizes a specific sequence.
- Southern hybridization can be used, but polymerase chain reaction (PCR) is now more frequently employed to obtain restriction fragment length polymorphisms (RFLP) or restriction phenotype patterns.
- PCR is a sensitive *in vitro* method for amplifying specific DNA sequences using two single-stranded oligonucleotide primers that anneal at each end of the segment to be amplified.
- Primers are the key to PCR because they direct DNA synthesis to particular segments.

- If both sequence and primers are available, PCR is the method of choice for detecting genotypic polymorphisms.
- Random amplification of polymorphic DNA (RAPD) is based on the fact that any random sequence of bases is expected to be complementary to some part of a target genome.
- RAPD-PCR can detect variation between fungal isolates not distinguished by RFLP because a single base substitution can prevent PCR primers annealing.
- Single-strand conformation polymorphisms (SSCP) can detect any base pair variant in a DNA sequence without prior knowledge of the sequence.
- DNA fingerprinting is the simultaneous detection of genotypic differences in a panel of unlinked polymorphic loci dispersed throughout the genome.
- DNA arrays on microchips offer the prospect of detection of genotypes rapidly, effectively, and automatically.
- DNA fingerprinting has been used in population studies because the technique can distinguish individuals.
- Microsatellites are small, extremely polymorphic loci also called simple sequence repeats (SSRs) and consisting of tandem repeats made up of a few bases.
- Microsatellite polymorphism is based on the number of reiterations of the repeated sequence.
- Minisatellites are tandem repeats 10–40 bp in length and repeated 10 to a few thousand times.
- Minisatellites are located near the ends of chromosomes; their polymorphism arises from misalignment during recombination.
- Transposable elements are also repeated and include transposons and retrotransposons.
- Retrotransposons in yeast are called Ty elements and are integrated into DNA using a proteinase, reverse transcriptase, and integrase.
- Other fungal transposons include *copia*- and *gypsy*-like elements in *Neurospora crassa* and MAgnaporthe Grisea GypsY (MAGGY) elements in *Magnaporthe grisea.*
- Pseudogenes are duplicate copies of functional genes that do not contain introns and are assumed to have arisen from reverse transcription from mRNA.
- Pseudogenes can also arise from gene duplication, but are inactive as a result of an accumulation of deleterious mutations that block transcription.
- Spacer DNA separates one gene from the next, or one gene cluster from the next, and provides a source of DNA polymorphisms.
- Nontranscribed spacers between ribosomal DNA transcription units have been extensively used as molecular markers.
- Karyotyping is the assessment of variation in chromosome number and size.

- Orthogonal-field alternating gel electrophoresis (OFAGE) and pulsed-field gel electrophoresis (PFGE) are methods used to separate chromosomes and detect chromosome length polymorphisms (CLPs).

7.1 Molecular Markers

The genetic analysis we have described so far has dealt with functional genes. In these, the detection of a difference in genotype depends on the functional expression of the genes concerned. Recognizing a white-spored mutation as a character distinct from the black spore phenotype obviously depends upon the function of the gene or genes that control pigmentation, but it is also dependent on the expression of all those other gene functions that contribute to sporulation. Unless the culture can be encouraged to sporulate, the spore color cannot be scored. Even though the mutated pigmentation gene is present in every nucleus its presence can only be scored in the specific cell type in which it is expressed. The fact remains, however, that the pigmentation mutation is represented in the genotype of the organism at the DNA level. The white-spored DNA must have a different sequence to the black-spored DNA. If that difference in DNA sequence could be detected directly using recombinant DNA technologies, then the resultant molecular markers would be scorable in DNA from any nucleus of the organism, quite independently of the functioning of the pigmentation gene.

Before DNA markers came to prominence, electrophoresis of proteins was a technique used to identify loci corresponding to molecular features that could be used for characterizing strains. Proteins can be separated on polyacrylamide and starch gels under conditions that preserve their enzymic activity. As long as a selective staining reaction can be devised, the pattern of bands on the electrophoretic gel due to a particular enzyme activity can be recognized against the background of all the other proteins in the extract. If the protein bands are variable, due to slight differences in the structures of the proteins between strains causing differences in electrophoretic mobility, then the technique allows identification of alleles based on the genetic determination of protein structure. These produce enzyme or protein polymorphisms, called *isozymes* (the multiple forms of an enzyme that arise from a genetically determined difference in primary structure) or *allozymes* (enzyme variants produced by particular alleles).

The technique is still useful, but protein polymorphisms compare unfavorably with DNA polymorphisms because protein electrophoresis assays the genotype indirectly, and a high proportion of the variation that occurs at the DNA level may be undetectable because it does not alter the amino acid composition of the protein. Some changes in amino acid composition similarly do not change electrophoretic mobility of the protein, and remain undetected, leading to different genotypes being assigned to the same

allozyme allele. Finally, because the technique deals with the gene products, allozyme alleles may be subject to selection and give a different view of population structure from that obtained with DNA markers. This is because so many of the latter are neutral and not under selection.

7.2 DNA Polymorphisms

Differences in DNA sequence are called *DNA polymorphisms* and they provide the basis for the direct physical analysis of the genotype using molecular methods. Some of the DNA sequence polymorphisms, like our spore color example, occur within functional genes. DNA polymorphisms, however, have two enormous advantages over conventional functional mutations. The first is that the sequence difference is detected directly and no functional phenotype need ever be associated with that sequence. DNA polymorphisms that have no known function are called *anonymous loci*. The second advantage of DNA polymorphisms is that they occur in a genome at very high frequency. One reason for their high frequency is that although functional gene mutations are by definition limited to coding regions, DNA polymorphisms can occur in any DNA sequence, whether it contributes to a coding region or not.

Up to 30% of an "average" segment of fungal nuclear DNA may be comprised of noncoding sequences (i.e., spacers, introns, and various sorts of repeated sequence). Variations in any of these regions are potentially detectable as DNA polymorphisms, but not detectable at all as functional mutations. It is worth noting that fungi tend to have fairly compact genomes. Although 30% noncoding may sound rather wasteful, other eukaryotes have very much more noncoding DNA; in humans and other higher eukaryotes, about 95% of the DNA can be comprised of noncoding sequence. It contributes to chromosome structure (e.g., the DNA of telomeres and centromeres), but otherwise has no known function.

Because such a small fraction of the DNA is involved in functional coding, most of these nucleotide changes do not have any effect on the phenotype of the organism. Having no effect on phenotype means that DNA polymorphisms are not subject to selection. They can persist in a population or disappear by chance. Because mutations in functional genes or their regulators are likely to be harmful and quickly eliminated from the gene pool by natural selection, the overwhelming majority of DNA variation that remains in a natural population will be that which is located in noncoding regions of DNA. As a "rule of thumb," sequence variations that occur at a frequency of 1% or more in a population are considered to represent polymorphic genetic loci, and it seems likely that most of these will be DNA polymorphisms.

Because of their larger and less compact genomes, DNA sequence polymorphisms are even more readily detectable. They have reached even

greater significance in animals and plants than they have in fungi, but there is another point worth mentioning that applies equally to all Kingdoms. The genetic code is redundant; the 64 codons of the genetic code cluster into groups, with each group coding for the same amino acid. Indeed, for 32 of the 64 codons the identity of the third base is irrelevant to the meaning of the codon, so a mutation in any of those positions will not result in any change of amino acid. Such mutations, however, will be potentially detectable as a change in sequence DNA polymorphism.

Observations on a wide range of organisms suggest that about one nucleotide difference every 700 bp would not be unusual, and that those differences are likely to be observed whether the comparison is made between two different individuals in a population or even between the two homologous chromosomes within one diploid. That level of difference corresponds to more than 3 million differences between any two haploid sets of human chromosomes and is obviously an enormous pool of potential variation in humans. The smaller genome size and lesser proportion of noncoding DNA in fungi result in a smaller pool of DNA markers, but we estimate that two unrelated yeast cells could differ from one another in as many as 3000 DNA polymorphisms!

7.3 Restriction Fragment Length Polymorphisms

DNA sequence differences arise through a variety of mutational events: transitions, transversions, deletions or insertions, and as a result of errors in replication (described in Chapter 3). The simplest DNA polymorphisms are changes in single bases. These are also the most numerous as they can make up more than 95% of all DNA polymorphisms. The simplest ones to detect are those that eliminate or create a recognition site for a restriction enzyme. These give rise to restriction fragment length polymorphisms (RFLPs), which were the first sort of DNA marker to be used. Restriction fragments are regularly sized pieces of DNA that are produced when a DNA sample is digested with a restriction endonuclease. Restriction endonucleases recognize specific DNA sequences, usually four to six nucleotides in length, and cut the DNA in or near the recognition sequence. For example, the archetypal restriction endonuclease is called *Eco*RI. It is obtained from a strain of *Escherichia coli* (giving the *Eco* part of the name) carrying a particular plasmid (that's the "RI" part of the name of the enzyme). *Eco*RI cuts double-stranded DNA only at the sequence:

5′-G↓AATTC

3′-CTTAA↓G

The cuts are made between the G and A, where we've put the vertical arrows. Digestion with this enzyme produces DNA fragments that all have

single-stranded ends with the sequence TTAA and which are capable of reannealing with each other (or with any other fragments produced by digestion with *Eco*RI). That is why this particular enzyme has been so widely used in recombinant DNA experimentation. The significant fact for our present discussion, however, is that the sequence specificity for digestion by a restriction enzyme such as this means that treatment of any sample of DNA with the enzyme should always produce the same set of fragments.

In a real sense, therefore, the enzyme produces a restriction phenotype of the DNA of a particular strain. It is this feature that enables restriction fragments to be used in genetic analysis. The restriction fragments can be visualized by Southern hybridization in which the DNA fragments that result from the restriction digestion are separated using agarose gel electrophoresis. The fragments of interest (established in earlier experiments) are detected by hybridization with probes labeled with radioactive or fluorescent markers.

The polymerase chain reaction (PCR) is more effective than Southern analysis and is more often used now. In this approach the DNA containing the RFLP is amplified by PCR to the extent that it can be visualized on the agarose gel by staining with an unselective fluorochrome like ethidium bromide. By applying a standard digestion and detection technique it is possible to obtain a restriction phenotype characteristic of any particular sample of DNA. The relevance of this to RFLPs is that DNA polymorphisms can alter the restriction phenotype by changing the pattern of recognition sites for the restriction endonuclease. Alteration of the endonuclease recognition sequence by nucleotide substitution, insertion, or deletion can prevent the restriction endonuclease from acting and change the fragments produced when the DNA is digested. For some restriction enzymes, methylation of nucleotides has the same effect, although DNA methylation is consistently low in the fungi as compared with plants and animals.

For example, suppose that we are interested in a 10 kb fragment of DNA that is defined in the first population we examine by having a GAATTC recognition site for the *Eco*RI restriction enzyme at each end. The restriction phenotype of an electrophoretogram of DNA extracted from a (haploid) member of population 1 will, consequently, feature a band migrating in the 10 kb position (Fig. 7.1). Now, suppose that there is a sequence about a thousand bases along this fragment which appears as AAATTC in population 1, but has mutated in a different isolate of the organism (population 2) to the GAATTC *Eco*RI restriction sequence. When DNA from this second population is digested and electrophoresed there will be no 10 kb band; instead, the electrophoretogram will have two bands of 2 kb and 8 kb. These *are* the RFLPs we have been talking about: the restriction nuclease digestions reveal a single 10 kb restriction fragment in one strain, but a pair of restriction fragments (2 kb and 8 kb) in the second strain. We may never know what genetic function (if any) is expressed by this region of the genome; the function does not matter because the genetically useful

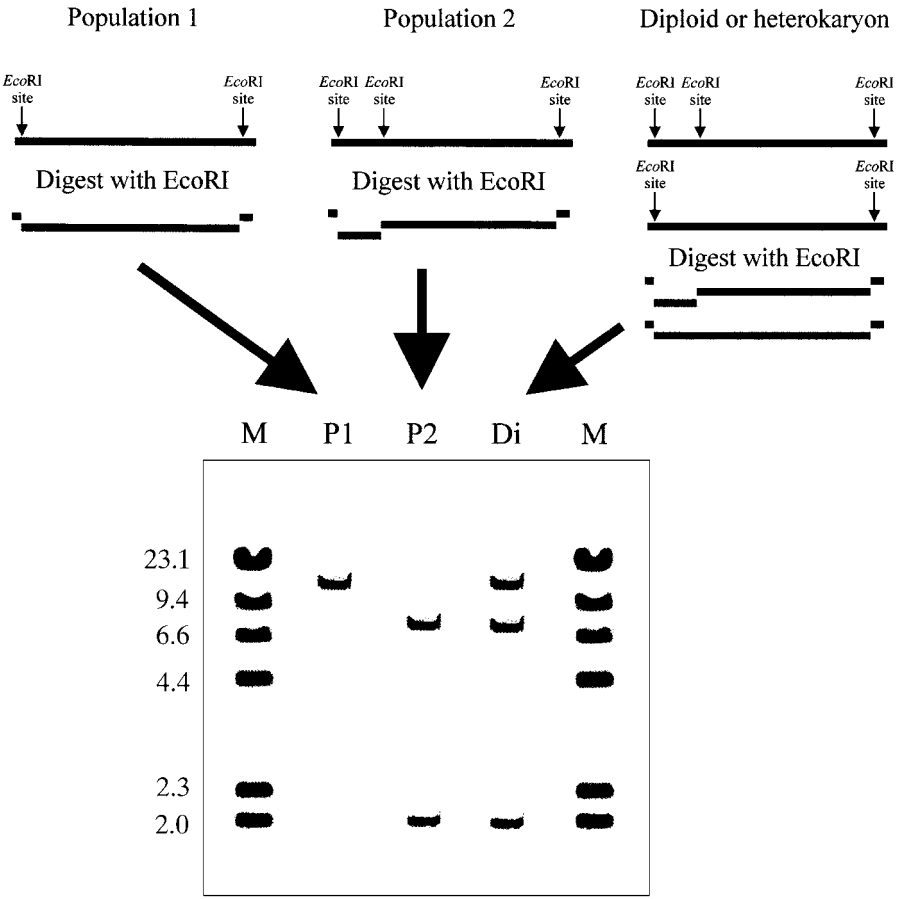


Fig. 7.1. Restriction phenotypes revealed on a DNA electrophoretogram. DNA extracted from a (haploid) member of population 1 features a band migrating in the 10 kb position. In population 2 a mutation has produced an additional GAATTC *Eco*RI restriction sequence. Digestion of this DNA produces two bands at 8 and 2 kb. This is the restriction phenotype that enables the two strains to be distinguished. DNA from a heterokaryon made between the two strains (or a dikaryon or diploid) will produce all three restriction fragments after digestion with *Eco*RI. This shows that the mutations are codominant; that is, the heterozygote produces both phenotypic characters. In the electrophoretogram represented at the bottom of the figure, the lanes labeled "M" were loaded with molecular weight markers, in this case DNA of the bacteriophage λ digested with the restriction enzyme *Hin*dIII. In each experimental lane, the small fragments of DNA indicated as arising from *Eco*RI digestion of the ends of the starting DNA would migrate rapidly through the gel and dissipate into the buffer.

phenotype is the number and size of the fragments produced by a standardized enzymic digestion of the DNA. This defines the two alleles of the RFLP we have discovered; there are only two alleles because we are scoring fragmentation by the restriction enzyme, and the two alternatives are the fragments that have the recognition site of the restriction enzyme, contrasted with fragments that do not have that recognition site.

When the two strains we have distinguished are compatible, we can make a heterokaryon, perhaps a dikaryon, carrying other nuclear types, and DNA extracted from the heterokaryon will contain all three restriction fragments: 10, 8, and 2 kb (Fig. 7.1). Mutations that disrupt or create recognition sites for restriction nucleases evidently generate easily detectable changes in the size of restriction fragments, and deletion or insertion mutations in the region between restriction sites also change the size of fragments in the gel pattern.

The ability to detect directly both alleles of an RFLP in DNA from the heterokaryon is taken to mean that the RFLP alleles, in common with other DNA markers, are codominant (i.e., equally expressed). Although the description is commonly used, however, dominance and recessiveness are aspects of the functional phenotype. The term *codominant* is not entirely appropriate because allelic differences in the structure of the DNA itself that are not expressed in the functional phenotype of the organism. RFLPs were the first DNA markers used to study fungal evolution, and they are still useful for this. There are, however, several independent ways in which a restriction endonuclease site can be changed due to the fact that any of the nucleotides in the recognition sequence can be substituted. Thus, completely unrelated isolates can suffer loss of the same site in different ways, and this can confuse the relationships on which evolutionary interpretations depend.

Nevertheless, RFLP alleles are ideal candidates for classical genetic (Mendelian) analysis because of the ability to detect the DNA variations of both alleles. Allele scoring is straightforward if there is only one variable restriction endonuclease site in the DNA fragment being detected. Interpretation can be difficult if there are several variable sites. It is better to construct a restriction map (see Section 8.3) of the fragment to aid interpretation of the segregation. Labeled probes can be used to visualize the fragments on Southern blots without any sequence information, so the RFLP locus can be mapped by following the inheritance of the alleles using the same approaches as are used when functional gene mutations are used as markers.

A six base pair recognition sequence (e.g., that of *Eco*RI) is expected to occur once every $4^6 = 4096$ bp in DNA, assuming random distribution of the bases. A genome of about 30 Mb, a size which is fairly representative of fungi, would therefore be expected to yield something like 7000 fragments when digested with *Eco*RI. Scoring a particular RFLP conse-

quently involves successfully probing for only one or two specific restriction fragments within a background of many fragments that are of no interest to us. The probe, which must usually be found by trial and error, is crucially important in this. It is not always easy to find RFLPs within regions of interest. Other disadvantages include the fact that this approach requires a relatively large amount of starting material as well as considerable time and effort in a procedure, including purification of DNA, restriction digestion of the DNA, running the digested DNA on a gel, making a Southern blot, preparing a suitably labeled probe, hybridizing probe to blot, and finally washing the blot to visualize the probe hybridization pattern.

7.4 Polymerase Chain Reaction

The PCR can be used to detect RFLPs that result from mutations in restriction enzyme sites. PCR is an *in vitro* method for amplifying specific DNA sequences that may be present in only trace amounts in a DNA sample from any source. PCR enzymatically generates millions or billions of exact copies and so makes genetic analysis of very small samples a relatively simple process. Two single-stranded oligonucleotide primers that anneal at each end of the segment define the segment that is amplified.

A typical PCR reaction is performed in a volume of 50–100 µl containing about 20 nmol of each of the four deoxynucleotide triphosphates (dATP, dCTP, dGTP, dTTP), 10–100 pmol of primers, magnesium salts and buffers, DNA polymerase, and the target DNA sample. The target DNA molecule is denatured by heating to about 94°C, then allowed to anneal with the primers by cooling to 50°C. A heat-stable DNA polymerase, (e.g., the *Taq* polymerase isolated from the extremely thermophilic bacterium *Thermus aquaticus*), is used to synthesize DNA complementary to the target by extension from the 3′-end of each primer at a temperature of about 72°C. Heating the reaction mixture to 94°C again detaches the new strands from the target DNA templates: one cycle of PCR doubles the number of target DNA molecules. When the mixture is cooled, more primers anneal to the target DNA *and* the newly synthesized strands, and a second cycle of DNA synthesis at 72°C takes place.

Primers that anneal to newly synthesized strands can only be extended to the position of the companion primer because, by definition, that's where synthesis of the newly synthesized strand was started. As the number of newly synthesized strands increases in successive cycles, more and more of them are made to exactly the size bounded by the 5′-ends of the two primers. The only molecules that are larger in size are the original target molecule and first generation complementary copies made on the original, in which extension from one primer can extend beyond the position to which the other would anneal.

Each extend–denature–reanneal cycle theoretically doubles the number of copies of the target sequence; this is a logarithmic rate of increase in product that implies that 20 cycles could theoretically result in a million-fold amplification of the DNA fragment. In practice, efficiency of amplification depends on the length of the segment amplified: most reactions aim to amplify sequences of less than 2 kb, although fragments up to 40 kb long can be amplified with some modification of the procedure. The PCR can usually be repeated for 30 or 40 cycles, although in later cycles efficiency reduces as reactants approach exhaustion and accumulation of the amplified product inhibits subsequent polymerization reactions.

If a restriction enzyme site exists in the amplified segment it can be demonstrated by digesting a sample of the PCR product before running it out on an agarose gel. The undigested PCR product will produce one band; the digested sample will produce two bands (assuming there is only one restriction enzyme recognition site in the PCR fragment). It is not necessary to use Southern blotting because PCR amplifies the fragment so much that the bands can be visualized directly in the gel. This makes the procedure much simpler and faster than traditional RFLP analysis. Application of the same primers and enzyme digestion to DNA samples from a variety of isolates will often identify restriction enzyme sites that are polymorphic in the populations tested.

7.5 PCR Primers: AP-PCR and RAPD

The primers are the key to PCR because they direct the DNA synthesis cycles to a particular segment of the target DNA. Spurious fragments can be amplified by false priming at sites within the target. To guarantee maximum yield of the correct target sequence, optimum reaction conditions for each pair of primers must be determined empirically by adjusting the magnesium, primer, and enzyme concentrations and altering the temperature cycling. If possible, complementarity between primer and the body of the target sequence should be avoided to prevent amplification of primer–primer concatenates in preference to the target sequence.

The main attraction of PCR is probably the extreme sensitivity of the technique, but this can also be a problem because it is highly susceptible to contamination from other samples being processed at the same time, from DNA contamination from the experimenter, from the environment, or from previously amplified PCR products. Positive and negative controls are essential to monitor contamination problems, and great care must be exercised during sample preparation and PCR reaction. If sequence data for a particular gene exists, PCR effectively allows large quantities of the gene to be synthesized *in vitro* using primers designed from that predetermined sequence. Computer programs are available that aid primer design, and automated synthesizers can then synthesize the primers.

If restriction sites are incorporated into the primer sequences, then the PCR product can be cloned into a suitable vector for further recombinant DNA manipulations. The predetermined gene sequence may have been sequenced directly from the genome of the organism of interest. Although there is a considerable cost in terms of time and effort in determining a sequence, PCR techniques become the method of choice for detecting genotype polymorphisms once both sequence and primers are available.

On the other hand, the sequence may have been obtained from one of the sequence databases on the Internet. The choice of which sequence to use might be made on the basis of some prior interest (e.g., "I want to clone the NADP-linked glutamate dehydrogenase from my fungus, so I might as well use the *Neurospora crassa* sequence as a starting point") or could be made for a more imaginative reason (e.g., "I wonder if my fungus has any sequences similar to the genes that code for rattlesnake venom?").

Knowing a sequence, one way or another, allows you to target a very particular, and perhaps even understood, part of the genome; however, it's not necessary to have any sequence information to devise primer sets. You can use arbitrary primers (for AP-PCR), which are simply chosen because they are available; either they are the ones you can scrounge from your colleagues or they're the cheapest you can buy. The point is, they are primers designed for other purposes and you "arbitrarily" choose to use them as sequences that might define any old fragment in the genome of your fungus.

The ultimate extension of this approach is to use random amplification to search for polymorphic DNA in a genome of which you have little or no prior knowledge. Random amplification of polymorphic DNA (abbreviated to RAPD, which is pronounced "rapid") is based on the fact that any random sequence of bases, about 10 bases in length, is expected to be complementary purely by chance to some part of a target genome. If two random sequences complementary to the random primer occur within a reasonable distance of each other on opposite strands of the DNA, PCR will amplify the region between them. A random 10-base sequence is expected to occur every 4^{10} (= 1,048,576) bases and a random 9-base sequence every 4^9 (= 262,144) bases. This approach is similar to RFLP analysis in that it assays DNA sequence variation in short regions. Rather than depending on changes in restriction endonuclease recognition sequences, however, RAPD and AP-PCR rely on the PCR primer sites. Nucleotide substitutions in one of the regions to which the primers are complementary, particularly the 3′-ends, can prevent primer annealing and PCR amplification.

The basic technique is to use a short primer to amplify RAPD loci from different isolates of your fungus. Suppose that in the reference isolate the primers amplify two loci (because the genome contains two pairs of sequences complementary to the primer that are sufficiently close together to be amplified). Now assume that in an isolate of the fungus obtained from

a different place one of the primer binding sites has had a single base substitution. This will prevent primer binding at one end of one locus and, consequently, prevent amplification of that locus. Gel electrophoresis of PCR products from both isolates, using ethidium bromide to visualize the bands, will reveal the polymorphism as a double band in the reference isolate lane, but only a single band in the lane for the second isolate.

RAPD-PCR can detect variation between isolates that are not distinguished by RFLPs. The procedure works because a single nucleotide substitution can prevent priming. Such exquisite sensitivity is bound to be responsive to even small differences in other aspects of the PCR technique; therefore, extremely strict experimental technique is a necessity. Reproducibility of RAPDs can be a problem. Even if reproducibility is assured, the problem remains that PCR products that have the same electrophoretic mobility do not necessarily have the same DNA sequences. More than 10% of RAPD bands with equal mobility in plants do not have homologous sequences. In addition, as is the case with RFLPs, there are several ways in which a primer site can be lost, so two isolates that lack the same band may not be homologous.

7.6 Single-Strand Conformation Polymorphisms

The solution to most of these problems is to sequence the RAPD bands to confirm their identity. Single-strand conformation polymorphisms (SSCP) can detect any base pair variant in a DNA sequence, without prior sequence information. The technique depends on the fact that single nucleotide changes can alter the three-dimensional conformation of single-stranded DNA and alter its electrophoretic mobility. As an example, take a PCR-amplified product, denature it at a high temperature, and then "snap-cool" it by placing it on ice. The rapid exposure to low temperature prevents reformation of double helixes, but the single strands fold into their most stable conformation, with hydrogen bonding occurring between the many bases that come together when the single strand folds. The hydrogen bonding stabilizes the conformation so that the molecule has a particular electrophoretic mobility.

In a polymorphic allele, substitution of one nucleotide for another within the strand will alter the stable conformation of that strand by changing the hydrogen-bonding pattern. The new conformation may have a different mobility on the gel. Thus, different PCR-amplified alleles that differ by single base substitutions will be detected because of the variations in the electrophoretic mobilities of their single strands; these are the single-strand conformation polymorphisms (SSCPs) that provide direct evidence of mutations within the selected sequence, as well as a means of following them as they segregate in crosses. The variable bands can be retrieved from

the gel, sequenced, and more specific PCR primers designed to amplify the particular polymorphic fragments of the alleles, which is a technique called sequence-confirmed amplified region (SCAR) analysis.

7.7 DNA Fingerprinting

Sequencing is being used much more frequently, but in the standard RAPD analysis you would use several to many random oligonucleotide primers separately to amplify random loci from a collection of isolates. This would give you a panel of polymorphic loci, with each one being detected as a DNA band present in the gel-lanes of some isolates but not in others. If the panel of polymorphisms is large enough, it should be possible to define each isolate in terms of its own unique RAPD profile. By this stage, each RAPD profile has become an identifier for the isolate, and then you might think about applying your ability to identify the different isolates by relating RAPD profiles with geographic origin, or using them as markers in crosses, or as ways to identify the presence of your fungus in soil samples, or as a contaminant of foodstuff.

Genetic approaches, and particularly those using molecular markers, are really essential in population studies of fungi. It is the matter of identity that makes them essential. Those who study animals and plants rarely experience any problem in recognizing individuals, yet fungi can be both difficult to detect and difficult to identify. As a result, establishing even the most basic features of a population (e.g., the size, number and distribution of its members) can pose major problems with fungi.

Analysis using molecular markers can detect polymorphisms at multiple loci, and rather than attempting to assign a specific band in the electrophoretic gel to a specific locus, it provides a DNA fingerprint that identifies the whole genome based on simultaneous detection of those multiple polymorphisms. RFLP analysis was the original basis of DNA fingerprinting, but RAPDs and AP-PCR require much less starting material and can be used with organisms that have not yet been well characterized genetically. DNA fingerprinting is the simultaneous detection of genotypic differences in a panel of unlinked polymorphic loci dispersed throughout the genome.

For the future, DNA arrays on microchips promise the prospect of detecting genotypes rapidly, effectively, and largely automatically. DNA arrays are microchips, a few centimeters squared, that hold several hundred thousand different DNA sequences. The production methods, either "etching" with masks like a computer chip or "printing" in tiny droplets like an inkjet printer, result in a checkerboard pattern of microscopic blocks, each of which is coated with single-stranded DNA of a particular allele. The array is then challenged with a DNA sample labeled with a fluorescent dye and also denatured into single strands. A DNA strand in the sample that is com-

plementary to one on the chip will hybridize to that block on the chip. After excess DNA has been washed away, the chip can be analyzed with an automated fluorescence microscope. The microscope's computer software interprets the pattern of blocks, which detects complementary strands, and reports the genotype of the sample.

With current technology, a single 2.5-cm^2 chip can screen 400,000 sequences at one time. These can be used to probe cDNA (DNA made *in vitro* using mRNA as template) that might, for example, represent the mRNA content of cells or tissues under different conditions or at different developmental stages. The sequences in the array might alternatively represent DNA samples from genomes that represent polymorphic loci able to define strains, species, or genera and therefore be used to identify the organisms present in samples that contain mixed DNA.

In current practice, the DNA fingerprint consists of an autoradiograph displaying the pattern of different-sized fragments produced from the genome of interest. It is most useful to have about 10–20 biallelic polymorphisms per genome because this number is most likely to allow resolution of all the allelic bands on the gel, yet be large enough to provide a highly discriminating fingerprint. If a single polymorphism exists (presence versus absence of a band on the gel), then there is a probability of 0.5 that two strains will have the same allele; for two polymorphisms the probability of two strains having the same alleles is 0.25. The probability is calculated from the function 0.5^N, where N is the number of polymorphic alleles being scored. With five, this works out to 0.03125, with ten 0.00098 (= a chance of one in a thousand), and with 20 polymorphisms there is a probability of 0.00000095 (= a one in a million chance) that two isolates will have the same alleles at all 20. Thus, a fairly small number of biallelic polymorphisms is able discriminate individuals in quite large populations.

Because DNA fingerprints can identify individuals, they have many applications in mycology. A few examples will illustrate how these and some of the other molecular tools we have described can be used for identification and characterization of fungi in which individuality is otherwise extremely difficult to establish. What is probably the most remarkable example appeared in a report in the journal *Nature* in 1992 with the title, "The fungus *Armillaria bulbosa* is among the largest and oldest living organisms." *A. bulbosa* is a facultative pathogen of the roots of trees, which means that it is not limited to a parasitic mode of life, but can grow as a saprotroph in the soil and in dead wood. It is common in European and eastern North American mixed hardwood forests. *Armillaria* forms ropelike aggregations of hyphae, called rhizomorphs, which spread in the forest soil seeking new substrates to acquire. Rhizomorph growth rates can be in the region of 0.2 m per year, which implies that a clone could become fairly large, fairly quickly. Conventional attempts to determine the size of individual clones of basidiomycetes used somatic compatibility reactions and distribution of mating-type alleles, and suggested that some clones could be up to 500 m in diameter.

Neither of these methods, however, can distinguish a single large clone from a number of closely related individuals because neither samples enough genes in the genotype. To make this distinction, the new survey used RFLPs and RAPDs to examine isolates from fruit bodies and rhizomorphs along 1 km transects through a forest in northern Michigan. A 15-hectare area yielded isolates that had identical mating-type alleles and mitochondrial DNA restriction fragment patterns, and 20 RAPD and 27 RFLP fragments were also all identical in all isolates from that area. So much genetic homogeneity indicates that the samples came from a single genetic individual, which is a clone, which had reached its 15-hectare size through vegetative growth. The clone is more than 600 m across, so an age of about 1500 years can be estimated from the rhizomorph growth rate.

Armillaria rhizomorphs are known to withstand the extremely high temperatures caused by forest fires, so the clone presumably lived through many fires that regenerated its forest. The mass of the clone was estimated from the fresh weight of rhizomorphs in the soil and the mycelium of fine hyphae in the soil and wood that supports rhizomorph growth. This very conservative estimate worked out to about 100 tons, which is close to the mass of an adult blue whale! Thus, the combination of molecular markers used in this study identified a single individual of *Armillaria bulbosa* that, at the time of the survey, occupied a minimum of 15 hectares, weighed more than 100,000 kg, and had remained genetically stable for more than 1500 years. That is a pretty amazing description of a fungal individual, but in the year 2000 similar molecular approaches revealed an individual of the related *Armillaria ostoyae*, estimated to be 2400 years old, that extended over an area of 890 hectares of the Malheur National Forest in eastern Oregon.

Not all the fungi of the forest, however, are monstrous. Field study in a remote broadleaved forest in Shaanxi province in China examined the natural distribution of *Lentinula edodes*. Using somatic incompatibility reactions and DNA fingerprints generated by AP-PCR, it was found that the territory of *L. edodes* genets was not more than 20 cm in diameter, which is the maximum distance between adjacent isolates that were sampled. Every fruit body collected was a different genet, and two sets of fruit bodies of different genets were so close together that the two fruit bodies distorted each other during development into aggregate fruits. This is a very different lifestyle from *Armillaria* and seems to result from a reproductive strategy in *L. edodes* that depends on basidiospore dispersal because there are no rhizomorphs, strands, or even asexual spores to aid vegetative dispersal of *L. edodes*.

The importance of these observations is that *Lentinula edodes* is the black oak mushroom, known in Japan as shiitake and in China as shiang-gu. It has been cultivated in China for more than 800 years and today is the third most popular cultivated mushroom in the world. China produces more than 70% of the world crop, and outdoor cultivation on logs is still the main cul-

tivation method on small family farms in China. Knowing that individualism in *L. edodes* is based on a strong somatic incompatibility system and that individuals arrive on their substrate in nature as airborne basidiospores is relevant to the mushroom industry as it is currently practiced in China. A view of native population structure has some important consequences for farming practice; in particular, arguing in favor of indoor cultivation to avoid cross-contamination between wild and cultivated strains.

A DNA probe cloned from the ectomycorrhizal *Laccaria proxima* revealed unique RFLP patterns in all isolates tested and was able to discriminate between *Laccaria* and other ectomycorrhizal fungi because it does not hybridize with the DNA of other fungi. The probe can be used to identify *Laccaria* among the ectomycorrhizas of trees, and can also reveal the population dynamics of the species and isolates identified. A similar study with the truffle, *Tuber borchii*, which is also ectomycorrhizal, found a RAPD fragment that could identify *T. borchii* fruit bodies, mycelia, and mycorrhizas, allowing all stages in the life cycle of the fungus to be monitored. On a larger geographic scale, eight RFLP loci have been used to study the intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. Almost 800 isolates were sampled from China, Japan, North America, and Europe, and the RFLP fingerprint genotypes compared. The comparison indicated that both RFLP allele frequencies and DNA fingerprints of North American and European subpopulations were similar to each other, and more similar to Japan than to China. This implies that *C. parasitica* was introduced into North America from Japan rather than from China.

Away from the forests, RAPD markers are able to differentiate two phenotypically similar species of *Penicillium* (*P. camembertii* and *P. nalgiovense*) used in the food industry, but PCR-RFLP prepared from the ITS region of the rRNA genes and of the 5′-end of the β-tubulin gene provided easy diagnostic markers to distinguish the two cultures.

7.8 Microsatellites

Microsatellites are relatively small, extremely polymorphic loci that are easily identified and characterized with PCR. Also known as simple sequence repeats (SSRs), microsatellites are tandem repeats of sequences made up of a few bases. The polymorphism of microsatellites is in the number of reiterations they carry of the basic repeated sequence. Error in DNA replication seems to be the major mechanism that generates new polymorphic alleles. The DNA polymerase may misread the template during replication, which causes slippage synthesis of a second copy of the same short sequence. Because there can be several different length variants, microsatellites can be multiallelic.

Tandem repeats of microsatellites have no known function. They are particularly common in vertebrates, especially mammals; for example, there are something like 100,000 microsatellite loci of at least 10 different sorts in the human genome. They vary at a rate of one mutation per 1000 gametes in humans, which is low enough for them to be relatively stable markers in human genetics and excellent candidates for genetic fingerprinting. Typical microsatellites are amenable to PCR typing, and this has made them the most important class of polymorphic markers for linkage mapping in mice, humans, and other vertebrates.

Microsatellites are much less common in fungi, but are still useful. For example, a large compound microsatellite in the genome of *Ascochyta rabiei*, which is a pathogen of the chickpea, contained the penta- and decameric repeat units CATTT, CATTA, CATATCATTT, and TATTT. In one study, a high level of sequence variation was observed in this pathogen, and a new length allele derived from a cross, presumably arising from recombination, had an alteration in the copy number of the TATTT repeat, from 53 to 65 copies. Primers designed to flank microsatellite-rich regions in the genome of *Sphaeropsis sapinea*, which is an endophyte of *Pinus* spp., established 11 polymorphic markers that distinguished different morphological types of the fungus. Identifying the fungus is important because it can cause disease in trees subjected to stress, and a molecular probe would permit the organism to be detected before the onset of symptoms.

Polymorphic microsatellite loci are defined by the actual size of the locus (number of repeats). Distinguishing the polymorphic alleles requires that the specific sequences at either end of the microsatellite are established and then used to design PCR primers. The flanking primers can then be used to amplify the locus from genomic DNA of different isolates, and the molecular size of the PCR products compared by gel electrophoresis. A single pair of locus-specific flanking primers will be able to detect many different alleles because a good gel electrophoresis system will detect differences as small as one or two base pairs between PCR products up to several hundred base pairs in length. Microsatellites can be very polymorphic with several different alleles present in a population.

7.9 Minisatellites

Microsatellites merge into minisatellites, which are also tandemly repeated sequences; however, in minisatellites the sequence, which may be repeated from 10 to a few thousand times, is 10–40 bp in length. Again, mammalian genomes contain a large number of minisatellites that are also called variable number tandem repeats (VNTRs). VNTR loci in humans are sequences of 1–5 kilobases that consist of variable numbers of a repeating unit from 15 to 100 nucleotides long.

Minisatellite polymorphism generally arises as a result of misaligned recombination, when homologous chromosomes line up wrongly because of the numerous short repeats, and unequal recombination produces two new alleles, one with fewer and the other with more repeat units than the parental chromatids. This is not the case in *Botrytis cinerea*, however, which has a minisatellite with an AT-rich 37 bp repeat unit in the intron of the ATP synthase gene. It is found at only one locus in the genome, but the number of tandem repeats varies from 5 to 11 in different isolates of the fungus. Because no recombination has been observed between alleles of the flanking regions, the minisatellite is thought probably to mutate by slippage during replication. The repeat unit has not been found in any other genus yet, including *Sclerotinia*, which is the closest relative to *Botrytis*.

Minisatellites tend to be located more frequently near the ends of chromosomes, so they may not be as useful as microsatellites, which tend to have a more dispersed distribution. Another problem with minisatellite alleles is that they tend to be too long for easy PCR amplification because the repeat units themselves are relatively long. Instead, the procedure for detecting minisatellites is much more like RFLP analysis. Genomic DNA is digested with a restriction enzyme that cuts close to either end, but not within, the minisatellite. The restriction fragments are then separated on a gel, and transferred to a Southern blot, which is probed with a labeled DNA fragment from the minisatellite locus. The probed blot will display the pattern of differently sized fragments that represent all the loci of that minisatellite in the genome.

7.10 Transposable Elements

Other repetitive DNA elements within eukaryotic genomes are transposable elements. These include transposons, which have propagated within the genome and moved into new locations as DNA, and retrotransposons, which are similar to retroviruses and propagate from RNA using the enzyme reverse transcriptase. Retrovirallike retrotransposons include the Ty elements of yeast (about 6 kb in length with about 30 full-length copies in each genome), *copia* and *gypsy* elements of the fruit fly *Drosophila*, and elements in mammals called long interspersed elements (LINES), or short interspersed elements (SINES).

The main human LINE, called *L1*, is up to 6.4 kb in length, and our genome contains approximately 20,000 copies of it. The human genome also contains 300,000 copies of the main human SINE, which is 280 bases long; this is called the *Alu* repetitive sequence because it contains a single target site for the *Alu*1 restriction enzyme. The full *Alu* sequence is similar to that of an RNA molecule that is a component of the signal recognition particle

that targets proteins that containing signal sequences to the endoplasmic reticulum membrane. *Alu* sequences presumably originate as reverse transcripts of these RNA molecules. Together, the two retrotransposons, *Alu* and *L1*, make up about 7% of the human genome, whereas 12.5% of the *Drosophila melanogaster* genome is made up of about 80 different retrotransposons, each an average 5 kb long, and present in an average of 50 copies. The human and, especially, the *Drosophila* retrotransposons were, for the most part, the first to be characterized and have been studied in the greatest depth. As a result, they tend to be used as comparison sequences with the result that similar sequences (e.g., "copialike" elements) are often the first to be found in other organisms.

Along with the retrotransposon elements from *Drosophila*, the Ty retrotransposons of the yeast *Saccharomyces cerevisiae* have become model retrotransposons in many respects. There are several yeast transposons (Ty 15 an acronym of *T*ransposon in *y*east), which are very similar in overall structure and are identified as Ty1 to Ty4. They contain a 5–5.3 kb internal domain that encodes proteins with proteinase, reverse transcriptase, and integrase activities. The reverse transcriptase transcribes the Ty-RNA into DNA, which the integrase, an endonuclease, is responsible for inserting into the genome. The proteinase performs proteolytic processing, which is a prerequisite of reverse transcription and integration.

The main sequence is flanked by long terminal repeats (LTRs), 330–370 bp in length, in the same orientation at each end of the element. These LTRs differ in length and structure between the various classes of Ty elements. LTRs also occur more frequently, and separately, from the Ty elements. These solo LTRs are thought to be relics of recombination between LTRs of complete Ty elements that have transposed away from the site. The reverse transcriptase encoded by the retrotransposon suggests that they move around the genome via an RNA intermediate. For example, a yeast Ty1 element transposing from a plasmid to a yeast chromosome lost the intron in one of its genes, suggesting that the transposon sequence passed through an RNA intermediate during transposition.

The major classes of Ty elements in laboratory strains of yeast are Ty1 and Ty2. Laboratory strains frequently have 30–35 copies of Ty1 and 5–15 copies of Ty2, among which there is considerable polymorphism due to base substitutions. Ty elements are widespread in natural isolates of *S. cerevisiae*, and the Ty pattern of a strain tends to be quite stable. Other retrotransposons are similar to retroviruses and also end in LTR nucleotide sequences 10–200 bp long, which are repeated in the same orientation at both ends of the retrotransposon, and transcripts are sometimes enclosed in viruslike particles.

As with other sorts of repetitive DNA, transposable elements account for a very much smaller proportion of the genome in yeasts and other fungi. Nevertheless, they are there to be found. A cluster of three sequences similar to retrotransposons and a degenerate fragment from a LINE-like

retrotransposon have been found in the centromere region of linkage group VII of *Neurospora crassa*. The region also contains a full-length but non-mobile *copia*-like element, and adjacent DNA contains portions of a *gypsy*-like element and a sequence with similarity to the Ty3 transposon of *Saccharomyces cerevisiae*.

Only one active transposon is known in *Neurospora*, but inactive DNA sequences have been found that are representative of different transposon families, and these bear unmistakable signs of inactivation by repeat induced point mutation (RIP, see Section 10.8). Such relics of transposable elements in *N. crassa* seem to be severely degraded retroransposons. The characteristics and arrangement of these degenerate retrotransposons in *Neurospora* are more similar to centromeres of *Drosophila* than to those of most yeast. The genome of *Magnaporthe grisea*, which causes the rice blast disease, contains numerous sequences similar to the *gypsy* retrotransposon; they're called MAGGY (MAgnaporthe Grisea GypsY). MAGGY elements are scattered throughout the fungal genome, and may be involved in genomic rearrangements because they can be associated with abnormal segregations. Other repetitive DNA sequences are located close to MAGGYs, which indicates that repetitive DNA sequences tend to cluster in the *M. grisea* genome.

Transposable elements do not seem to be useful to the organism that harbors them, and they are often described as "selfish-DNA" entities that carry information only for their own self-perpetuation; however, they are extremely useful for DNA fingerprinting. A lot of what we know about transposable elements has come from work with *D. melanogaster* in which the genetics and cytogenetics (especially the giant polytene chromosomes) make detailed analysis much more purposeful. For example, a probe including the *copia* transposable element would hybridize to 30–50 sites scattered throughout the genome of *D. melanogaster*.

The polytene chromosomes of *D. melanogaster* enable easy *in situ* hybridization, and if applied to preparations from two strains of flies captured in different geographic locations, you would find that the chromosomal sites to which the probe hybridizes *in situ* would be different in the two strains. This is the indication that since the time when the two strains were geographically isolated, the *copia* sequences have moved around (transposed) in different ways in the two genomes. Many copies of transposable elements are defective as a result of deletions that occur during transposition. Some of these deletions remove sequences needed for transposition so that this particular copy of the transposon will then be immobile. Transposition can generate mutations in genes into which the transposon inserts or alongside which they insert. It can also generate chromosomal rearrangements and relocate genes. The impact of transposable elements on genetics goes right back to the origin of the science: the wrinkled pea phenotype studied by Gregor Mendel was caused by insertion of a transposable element into a gene for a starch-synthesizing enzyme.

7.11 Genes and Spacers

Processed pseudogenes are another moderately repetitive element that might be scattered in the genome. They are duplicate copies of functional gene sequences, but because they do not contain the introns characteristic of the functional gene, they are also assumed to have been created by reverse transcription from mRNA. They may be located on a different chromosome from the functional gene as a result of a DNA integration event following reverse transcription of the mRNA. Other pseudogenes arise as the result of duplication, so their DNA sequence has close similarity to a functional gene, including its introns. The pseudogene is not expressed as functional protein because of the accumulation of deleterious mutations that block transcription, prevent correct RNA splicing, or introduce premature termination codons.

Finally, spacer DNA, which separates one gene from the next in any gene cluster, or one gene cluster from the next, can provide a source of DNA polymorphisms. Most use has probably been made of the internally transcribed spacers (ITS) within the transcription units of the DNA that correspond to the ribosomal RNAs. The spacers contain the exonuclease cleavage sites that are cut to generate the mature rRNAs, but the ITS sequences are repetitive, polymorphic, and ideal candidates as molecular markers.

The nontranscribed spacers between the ribosomal DNA transcription units vary in length due to variation in number of an internally repetitious sequence, so these are also ideal candidates as molecular markers. Because protein synthesis is such an ancient process, ribosomal nucleotide sequences are functionally constant, universally distributed, and moderately well conserved; additionally, they have proved to be extremely useful in establishing evolutionary relationships. We will deal with this in detail in Chapter 9. Polymorphisms in ribosomal DNA sequences can also be used to detect isolates and indicate relatedness. For example, RFLP analysis of PCR-amplified ribosomal RNA intergenic spacer regions has been used to examine genetic relationships between 23 isolates of the insect pathogen *Paecilomyces farinosus* collected in eastern Canada. About 40% of the spacer regions were variable, and 17 of the *P. farinosus* isolates had identical or nearly identical RFLP patterns, although they had been collected from six different hosts in different habitats. Neither host nor habitat is seemingly directing the genetic variation of these populations.

Whereas the preceding case indicates that there may be a high degree of similarity in ribosomal DNA sequences among isolates, there are as many examples where there is a sufficient degree of uniqueness in ribosomal DNA to be able to identify species and differentiate strains as in the case of dermatophyte fungi. The polymorphisms were detected by hybridization of *Eco*RI-digested genomic DNA of the dermatophyte fungus *Trichophyton rubrum* with a probe PCR-amplified from the small-subunit (18S)

rDNA. These RFLPs mapped to the nontranscribed intergenic spacer (NTS) region of the rDNA repeat and were able to discriminate 14 RFLP patterns among 50 random clinical isolates of *T. rubrum*, although 38% of the isolates shared just one of these RFLP patterns. Use of PCR with universal ITS primers produced amplified ITS products, which proved to have unique and easily identifiable fragment patterns for a majority of species after digestion with the restriction endonuclease *Mva*I. Such a study shows that RFLP analysis of the NTS and ITS intergenic regions of the rDNA repeat enables both molecular strain typing of *T. rubrum* and species identification of other common dermatophyte fungi.

A similar approach has also been developed to differentiate wood-staining fungi. Several different genera of fungi are able to grow in or on wood, producing pigments that reduce the value of the timber. Early detection is important so that steps can be taken to control the infestation, ideally before the timber is stained by extensive growth of the fungus. PCR is ideal for the detection of all sorts of pathogens because it is possible to amplify a specific sequence present at low concentration within a complex DNA mixture. Thus, it was used to develop a DNA fingerprinting method for sapstain fungi by amplifying small subunit ribosomal DNA (SSrDNA) from 55 fungal isolates of 13 sapstain species belonging to the genera *Aureobasidium*, *Ceratocystis*, *Leptographium*, and *Ophiostoma*. The amplified SSrDNAs were digested with 10 restriction enzymes to find polymorphisms useful for differentiating the isolates. Genus-specific RFLPs for all four genera were defined by the restriction enzymes *Rsa*I, *Sty*I, and *Taq*I, providing very efficient tools for identification of the major sapstain fungi on stained wood.

7.12 Electrophoretic Karyotypes

Assessment of the variation in chromosome number and size, called *karyotyping*, uses conventional microscopic techniques and is not feasible with fungi because of the small size of fungal chromosomes. It is fortunate that a novel electrophoresis system was introduced in 1984 that was able to separate DNA molecules of chromosomal size (i.e., up to 2 million bases in size, which is two megabases or 2 Mb). The technique uses two alternating electric fields and was called orthogonal-field alternating gel electrophoresis (OFAGE). The alternating electric field forces the DNA molecules to reorient at every change of field, and because reorientation time depends on the size of the DNA molecule, very large molecules are separated from one another in a size-dependent manner. A number of different systems have been described, and most use is now made of a process called pulsed-field gel electrophoresis (PFGE), which can separate molecules up to 10 Mb. Karyotyping based on PFGE reveals that both chromosome

number and size can be highly variable in fungi and can be used for isolate identification.

Prerequisites for the successful separation of fungal chromosomes are preparation of an adequate quantity of intact protoplasts, which are then embedded into the agarose gel, and appropriate electrophoresis conditions to separate the chromosomes. Fungal karyotypes can generally be divided into the two groups: "*Saccharomyces*-like," with chromosomes in the range 200 kb to 2 Mb that are electrophoretically separated by pulse times varying from one to several minutes in a total run time of 20–24 h, and "*Schizosaccharomyces*-like," with larger chromosomes 2–8 Mb in size, that require both longer pulse times of 30–60 minutes and run times that can extend to 5–8 days. Many filamentous fungi belong to this second group, although plant pathogenic fungi tend to have smaller chromosomes. Fig. 7.2 shows electrophoretic karyotypes of some isolates of *Aspergillus nidulans*. *A. nidulans* has eight chromosomes, which resolve into six bands between 5 and 2.9 Mb in size (chromosomes I and V are the same size and run as a single band, as do III and VI). The electrophoretogram shown includes a lane loaded with the much smaller chromosomes of *Saccharomyces cerevisiae* and illustrates how these run much faster in the gel under conditions that nicely separate the *A. nidulans* chromosomes. Comparison of several wild isolates of *A. nidulans* reveals relatively small length variations between their chromosomes, although many other fungi, including *A. fumigatus*, do contain a great deal of chromosome-size variation in natural populations.

The location of individual "genes" can be established by using either known gene sequences or RFLP or RAPD sequences to probe the gel bands produced by electrophoretic karyotyping. Individual chromosome bands can alternatively be cut out of the PFGE gel and the DNA digested with endonucleases to produce chromosome-specific clones. PFGE is also used in the construction of long-range restriction maps. Most restriction endonucleases cut chromosomal DNA into small fragments, but there are a number of so-called rare cutter enzymes, which produce much larger fragments because they have recognition sites more than six base pairs in length or a base composition very different from that of the target DNA. The (large) restriction fragments can be separated by PFGE and restriction maps can be produced from individual chromosomes as a prelude to mapping fungal genomes.

Chromosome length polymorphisms (CLPs) are detectable by electrophoretic karyotyping and can differentiate between genetically distinct organisms. Even closely related strains exhibit CLPs and they have been observed in many fungi, including *Saccharomyces cerevisiae*, *Candida albicans*, *Colletotrichum gloeosporioides*, *Aspergillus nidulans* (Fig. 7.2), *Acremonium chrysogenum*, *Septoria tritici*, and *Schizophyllum commune*. CLPs can arise from chromosome rearrangements like translocations. Pathogenic fungi may have dispensable or supernumerary B chromosomes. These

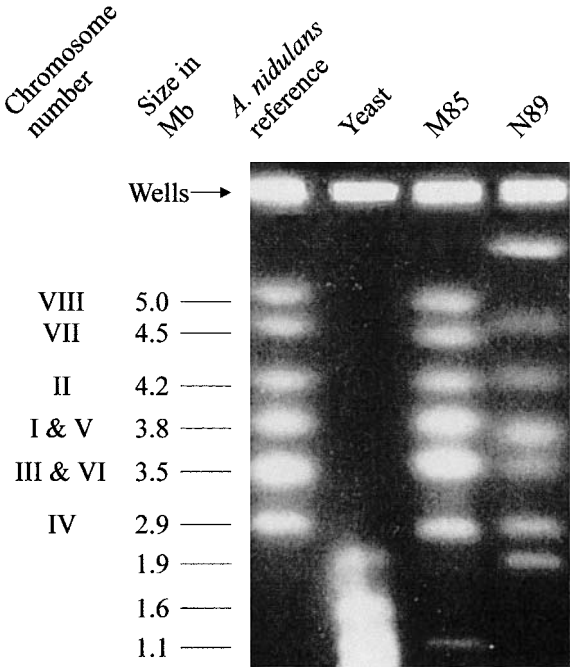


Fig. 7.2. Electrophoretic karyotypes of some isolates of *Aspergillus nidulans*. The *A. nidulans* reference strain is a "wild type" commonly used in laboratory study. It has eight chromosomes, which resolve into six bands between 5 and 2.9 Mb in size (chromosomes I and V are the same size and run as a single band, as do III and VI). Chromosomes of yeast (*Saccharomyces cerevisiae*) were run in the second lane and are much smaller than the *A. nidulans* chromosomes, and consequently run much faster in this gel. M85 and N89 are wild isolates that have unusual karyotypes. M85 has a B-chromosome (about 1.1 Mb in size), which is evidently dispensable because other strains do not have it, and derived strains show no phenotype when B-chromosome is lost. N89 has one unusually large and one unusually small chromosome resulting from a nonreciprocal translocation of about 1.6 Mb from the right arm of chromosome VI onto chromosome VIII. Comparison of several other wild isolates revealed relatively small-length variations between their chromosomes, although many other fungi do contain a great deal of chromosome-sized variation. (Based, with permission, on an illustration in Geiser, Arnold & Timberlake (1996), *Current Genetics* **29**, 293–300.)

tend to be small in size, generally do not contain functional genes, and are unstable during meiosis.

Publications and Websites Worth a Visit

Bertini, L., Agostini, D., Potenza, L., Rossi, I., Zeppa, S., Zambonelli, A. & Stocchi, V. (1998). Molecular markers for the identification of the ectomycorrhizal fungus *Tuber borchii*. *New Phytologist* **139**, 565–570.

Bowyer, P. (2001). DNA-mediated transformation of fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 33–46. Oxford University Press: Oxford, UK.

Burgess, T. Wingfield, M.J. & Wingfield, B.W. (2001). Simple sequence repeat markers distinguish among morphotypes of *Sphaeropsis sapinea*. *Applied and Environmental Microbiology* **67**, 354–362.

Cambareri, E.B., Aisner, R. & Carbon, J. (1998). Structure of the chromosome VII centromere region in *Neurospora crassa*: degenerate transposons and simple repeats. *Molecular and Cellular Biology* **18**, 5465–5477.

Chew, J.S.K., Strongman, D.B. & MacKay, R. (1997). RFLP analysis of rRNA intergenic spacer regions of 23 isolates of the entomopathogen *Paecilomyces farinosus*. *Canadian Journal of Botany* **75**, 2038–2044.

Chiu, S.W., Wang, Z.M., Chiu, W.T., Lin, F.C. & Moore, D. (1999). An integrated study of individualism in *Lentinula edodes* in nature and its implication for cultivation strategy. *Mycological Research* **103**, 651–660.

Debets, F. (1996). Use of a heterologous gene as marker for genetic analysis in *Aspergillus nidulans*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 273–280. Marcel Dekker, Inc.: New York.

Dupont, J., Magnin, S., Marti, A. & Brousse, M. (1999). Molecular tools for identification of *Penicillium* starter cultures used in the food industry. *International Journal of Food Microbiology* **49**, 109–118.

Geiser, D.M., Arnold, M.L. & Timberlake, W.E. (1996). Wild chromosome variants in *Aspergillus nidulans*. *Current Genetics* **29**, 293–300.

Geistlinger, J., Weising, K., Kaiser, W.J. & Kahl, G. (1997). Allelic variation at a hypervariable compound microsatellite locus in the ascomycete *Ascochyta rabiei*. *Molecular and General Genetics* **256**, 298–305.

Giraud, T., Fortini, D., Levis, C. & Brygoo, Y. (1998). The minisatellite, MSB1, in the fungus *Botrytis cinerea*, probably mutates by slippage. *Molecular Biology and Evolution* **15**, 1524–1531.

Goosen, T. & Debets, F. (1996). Molecular genetic analysis. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 97–117. Marcel Dekker, Inc.: New York.

Jackson, C.J., Barton, R.C. & Evans, E.G.V. (1999). Species identification and strain differentiation of dermatophyte fungi by analysis of ribosomal-DNA intergenic spacer regions. *Journal of Clinical Microbiology* **37**, 931–936.

Kim, S.H., Han, A., Kronstad, J. & Breuil, C. (1999). Differentiation of sapstain fungi by restriction fragment length polymorphism patterns in nuclear small subunit ribosomal DNA. *FEMS Microbiology Letters* **177**, 151–117.

Magee, P.T. & Scherer, S. (1996). Construction of a physical map of the *Candida albicans* genome. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 259–272. Marcel Dekker, Inc.: New York.

Milgroom, M.G., Wang, K.R., Zhou, Y., Lipari, S.E. & Kaneko, S. (1996). Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**, 179–190.

Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y. & Mayama, S. (1999). Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Molecular and General Genetics* **261**, 958–966.

Riach, M.B.R. & Kinghorn, J.R. (1996). Genetic transformation and vector developments in filamentous fungi. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 209–233. Marcel Dekker, Inc.: New York.

Smith, M.L., Bruhn, J.N. & Anderson, J.B. (1992). The fungus *Armillaria bulbosa* is among the largest and oldest living organisms. *Nature* **356**, 28–431.

Sweeney, M., Harmey, M.A. & Mitchell, D.T. (1996). Detection and identification of *Laccaria* species using a repeated DNA sequence from *Laccaria proxima*. *Mycological Research* **100**, 155–1521.

Talbot, N.J. (2001). Nucleic acid isolation and analysis. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 23–31. Oxford University Press: Oxford, U.K.

Taylor, J.W., Geiser, D.M., Burt, A. & Koufopanou, V. (1999). The evolutionary biology and population genetics underlying fungal strain typing. *Clinical Microbiology Reviews* **12**, 126–146.

von Ciriacy, M. (1995). Yeast retrotransposons. In *The Mycota*, vol. 2. *Genetics and Biotechnology* (U. Kück, ed.), pp. 227–245. Springer-Verlag: Heidelberg.

Walz, M. (1995). Electrophoretic karyotyping. In *The Mycota*, vol. 2. *Genetics and Biotechnology* (U. Kück, ed.), pp. 61–73. Springer-Verlag: Heidelberg.

Zhang, Y. & Molina, F.I. (1995). Strain typing of *Lentinula edodes* by random amplified polymorphism markers. *FEMS Microbiology Letters* **131**, 17–20.

Zolan, M.E. (1995). Chromosome-length polymorphism in fungi. *Microbiological Reviews* **59**, 686–698.

Zolan, M.E., Heyler, N.K. & Stassen, N.Y. (1994). Inheritance of chromosome-length polymorphisms in *Coprinus cinereus*. *Genetics* **137**, 87–94.

Historical Publications Worth Knowing About

Brody, H., Griffith, J., Cuticchia, A.J., Arnold, J. & Timberlake, W.E. (1991). Chromosome-specific recombinant DNA libraries from the fungus *Aspergillus nidulans*. *Nucleic Acids Research* **19**, 3105–3109.

Cooley, R.N. & Caten, C.E. (1991). Variation in electrophoretic karyotype between strains of *Septoria nodorum*. *Molecular and General Genetics* **228**, 17–23.

Fekete, C., Nagy, R. Debets, A.J.M. & Hornok, L. (1993). Electrophoretic karyotypes and gene mapping in eight species of the *Fusarium* sections Arthrosporiella and Sporotrichiella. *Current Genetics* **24**, 500–504.

Francis, D.M. & Michelmore, R.W. (1993). Two classes of chromosome-sized molecules are present in *Bremia lactucae*. *Experimental Mycology* **17**, 284–300.

Fukuda, M., Nakai, Y.F., Hibbett, D.S., Matsumoto, T. & Hayashi, Y. (1994). Mitochondrial DNA restriction fragment length polymorphisms in natural populations of *Lentinula edodes*. *Mycological Research* **98**, 169–175.

Khush, R.S., Becker, E. & Wach, M. (1992). DNA amplification polymorphisms of the cultivated mushroom *Agaricus bisporus*. *Applied and Environmental Microbiology* **58**, 2917–2977.

Kulkarni, R.K. (1991). DNA polymorphisms in *Lentinula edodes*, the shiitake mushroom. *Applied and Environmental Microbiology* **57**, 1735–1739.

Kwan, H.S., Chiu, S.W., Pang, K.W. & Cheng, S.C. (1992). Strain typing in *Lentinula edodes* by polymerase chain reaction. *Experimental Mycology* **16**, 163–166.

Lai, E., Birren, B.W., Clark, S.M., Simons, M.I. & Hood, L. (1989). Pulsed field gel electrophoresis. *BioTechniques* **7**, 34–42.

Lodder, S., Gull, K. & Wood, D.A. (1993). An electrophoretic karyotype of the cultivated mushroom—*Agaricus bisporus*. *Current Genetics* **24**, 496–499.

Loftus, M.G., Moore, D. & Elliott, T.J. (1988). DNA polymorphism in commercial and wild strains of the cultivated mushroom, *Agaricus bisporus*. *Theoretical and Applied Genetics* **76**, 712–718.

Magee, B.B., Koltin, Y., Gorman, J.A. & Magee, P.T. (1988). Assignment of cloned genes to the seven electrophoretically separated *Candida albicans* chromosomes. *Molecular and Cellular Biology* **8**, 4721–4726.

Plummer, K.M. & Howlett, B.J. (1993). Major chromosomal length polymorphisms are evident after meiosis in the phytopathogenic fungus *Leptosphaeria maculans*. *Current Genetics* **24**, 107–113.

Royer, J.C., Hintz, W.E., Kerrigan, R.W. & Horgen, P.A. (1992). Electrophoretic karyotype analysis of the button mushroom, *Agaricus bisporus*. *Genome* **35**, 694–698.

Schwartz, D.C. & Cantor, C.R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell* **37**, 67–75.

Talbot, N.J., Oliver, R.P. & Coddington, A. (1991). Pulsed field gel electrophoresis reveals chromosome length differences between strains of *Cladosporium fulvum* (syn. *Fulvia fulva*). *Molecular and General Genetics* **229**, 267–272.

Welsh, J. & McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**, 7213–7218.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A. & Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531–6536.

CHAPTER 8

Genes to Genomics: Mapping the Fungal Genome

Revision Concepts

- Physical maps compensate for the lack of resolution in genetic maps based solely on recombination frequencies.
- Mapping panels are DNA and/or culture samples used worldwide for linkage studies.
- Sequence tagged sites (STSs) can designate unique sequences, whereas expressed sequence tags (ESTs) can locate particular genes and provide information about their expression patterns.
- Ribosomal genes are popular for molecular analysis because they are repeated and their RNA products are abundant in the cytoplasm.
- Optical mapping involves restriction mapping individual DNA molecules without electrophoresis by using fluorescence light microscopy to examine the DNA.
- Cloning requires vectors that carry the DNA fragment of interest along with recognition sites for restriction enzymes; vectors differ widely in their size capacity for DNA: plasmids are the smallest vectors (up to 15 kb), whereas yeast artificial chromosomes (YACs) can carry up to 1 Mb.
- Chain termination sequencing, the method of choice, is based on enzymic synthesis of complementary DNA using radioactively or fluorescently labeled nucleotides, relying on primers to start and dideoxynucleotides to terminate the process.
- Alternative approaches of producing physical maps include pyrosequencing and DNA arrays.
- Strategies for compiling sequence information are shotgun sequencing, clone contigs (or chromosome walking), and directed sequencing.
- The complete genome (12.8 Mb comprising approximately 5570 genes) of *Saccharomyces cerevisiae* αS288C was sequenced from 1992 to 1996.
- The yeast genome has very few introns, but a high level of redundancy.
- Sequence comparisons have established surprising relationships between archaea and bacteria to explain eukaryote evolution.

- Model organisms under genomic scrutiny are the yeasts, *Candida albicans*, and *Schizosaccharomyces pombe*; clinical, plant pathogenic, and industrial filamentous fungi (e.g., *Aspergillus nidulans*, *A. fumigatus*, *A. niger*, *Pneumocystis carinii*, *Cryptococcus neoformans*, *Neurospora crassa*, *Magnaporthe grisea*, *Phytophthora sojae, P. infestans*, *Trichoderma reesei*); and several basidiomycetes (e.g., *Phanerochaete*, *Schizophyllum*, *Coprinus*, *Lentinula*, and *Agaricus*).

8.1 Genes and Maps: The Story So Far

Although we discussed molecular markers in the Chapter 7, and thereby leaned toward physical analysis, the theme of our story so far has been the use of the classic approach of genetics: making an organized series of crosses and using gene segregations to deduce chromosome maps. Entirely conventional procedures can result in those chromosomal maps showing the location of genes and centromeres. Then, by using molecular markers, the telomeres of the chromosome, and as many other loci as we are willing to characterize and include in our crosses, can be mapped along the chromosomes of an organism. Large-scale, high-density linkage maps can be assembled in this way. A high-density linkage map is conventionally considered to be one with one gene or marker for each 1% recombination in the genome.

In Chapter 4, we showed how two-point crosses can detect linkage between the two loci, and that the frequency of recombination provides a measure of the distance separating them. In Chapter 5, we introduced three-point crosses to integrate linked loci into a linkage map depicting both the order of loci as well as the distances between them on their chromosome. By widening the scope of loci to include those molecular markers that allow direct detection of genotype, like RFLPs, RAPDs, and micro- and mini-satellites, large numbers of loci can be scored in the progeny of a single cross, yielding sufficient data to build high-density linkage maps. Apart from the use of molecular techniques to determine the genotype of progeny, the analytical approach used with a multipoint cross is no different from that used with three-point crosses.

The reiterative calculations needed to analyze the accumulated data, however, have to be done by computer. Software capable of compiling data from 1000-point crosses has been developed for use with higher eukaryotes, so computer analysis of mapping data is not a limiting factor, although generating the data in the first place might be. Although the scale of analysis represented by 1000-point crosses is not necessary in fungi, largely because of their small chromosomes, the general approach is certainly applicable. In particular, researchers working toward linkage maps of mice and humans developed the concept, in the 1980s, of using mapping panels to simplify the problem of collecting large data sets together in a reasonable time.

The mapping panel is usually combined with a mapping program (e.g., MapMaker), which identifies linkage relationships, subject to the statistical constraints selected by the user.

A mapping panel consists of DNA samples and/or cultures from the same set of progeny that geneticists from laboratories around the world use for their linkage studies. A jointly characterized mapping panel allows easy introduction of new loci because so much is already known about the progeny in the panel that only genotype information for the new locus is needed. The new data can be added to a central database, and the computer program performs the linkage analysis. Thus, the number of loci scored in the progeny can increase, one or a few at a time, toward having hundreds to thousands of loci mapped. For example, in 1984 an RFLP mapping system for *Neurospora crassa* was established with two crosses involving multiple sequence differences. Progeny were widely distributed and are still used for mapping. The mapping panel from the first cross comprises 38 progeny from 18 ordered asci, and that from the second cross comprises 18 random ascospore progeny. The mapping panel from the first cross has been preferred over the years because it includes nonsister spores from the same half of ordered asci. As a result, distance from the centromere can be estimated, and double crossovers, gene conversions, and scoring errors are readily recognized. Updated maps for both crosses are regularly published in the *Fungal Genetics Newsletter*, which can be accessed through the Fungal Genetics Stock Center's Website.

The Human Genome Project has been at the forefront in the development of the technology and approaches used for physical mapping. An early aim of this project was to establish the locations of unique sequences, which are not duplicated at any other site, as markers spaced about 100 kb apart (that's just under 1% recombination) throughout the genome. These are called sequence tagged sites (STSs), and the most useful STSs in human genetics are highly polymorphic; therefore, there are enough different alleles to ensure that most people are likely to be heterozygous. The idea is that if the accumulated STS data were stored in Internet databases, anyone anywhere in the world would be able to use a bank of PCR primers to establish STS genotypes when mapping human genetic traits.

Microsatellites composed of repeats of two- or three-base sequences proved to be the most useful STS markers in humans because they usually have three, four, or even more alleles, so the probability of an individual being a heterozygote for a particular microsatellite locus is very high. The sequences surrounding microsatellite loci provide the positional uniqueness required for STSs. In humans, heterozygosity of a locus is a prerequisite for its inclusion in linkage analysis because you have to study pedigrees as genetic crosses are out of the question. This is not a consideration in experimental organisms, but the principle still applies, even to haploids. In all organisms, repetitive sequences provide good markers, and polymorphic markers are more likely to be found to be heterozygous in the parents of

any single cross that has been used to provide the progeny that make up the mapping panel.

Specific genes of interest can be targeted for inclusion in physical maps and/or sequencing projects by using expressed sequence tags (ESTs). These are a type of STS in which the sequence is defined by the cDNA (complementary DNA), which is synthesized from the messenger RNA molecules of a cell. The mRNA population of a cell represents the genes that are being expressed, and mRNA sequences lack the introns, spacers, and other noncoding sequences present in genomic DNA. Most eukaryotic mRNAs have a poly-A tail at their 3′-end, which can be made use of for their purification and then used as a primer site for converting the mRNA sequence into cDNA with reverse transcriptase. The collection of cDNA molecules can then be cloned so that genes of interest can be sequenced to a sufficient extent to provide probes for a mapping panel. Indeed, the same probes can also be used to identify the gene in genomic clones, and in these the introns and control sequences will be preserved. The population of mRNA molecules will be different at different stages of development and in different tissues. These differences are identifiable, and ESTs then provide a means to focus attention on genes with particular expression patterns.

Use of mapping panels is essentially a management technique for sharing the effort between participating laboratories, something that has been common for many years in several physical sciences, like particle physics and astronomy, although this represented the entry of biology into the "big-science" league. The first example of really big fungal science was initiated in 1989 when the European Commission initiated the program to sequence the yeast genome. The project involved 35 European laboratories at the outset, and the first sequence of a complete chromosome was published in 1992. More than 600 scientists were eventually involved, at locations in Europe, North America, and Japan, and progress involved distribution of DNA fragments to the contributing laboratories by the DNA coordinator. The complete sequence of the yeast genome was published in 1997.

Chromosome maps constructed solely from recombination frequencies have a limited resolution, although in microorganisms large numbers of progeny can be scored, thereby reducing this problem to some extent. When the *Saccharomyces cerevisiae* genome sequencing project began, the conventional genetic map consisted of more than 1400 markers, an average of 1 every 3.3 kb. This was detailed enough for the sequencing program without the need for much more physical mapping; however, *S. cerevisiae* is one of the two really intensively mapped eukaryotes (the fruit fly, *Drosophila*, is the other), so physical mapping will be necessary to improve the marker density in other fungi as the genome sequencing programs are expanded to include them. Physical maps are also needed to compensate for occasional inaccuracies in genetic maps based on recombination frequencies. Remember (from Chapters 4 and 5) that recombination maps

depend on the assumption that crossovers occur at random, but we can easily find evidence that such an assumption is only partially correct.

The sort of inaccuracy that can result was revealed when the complete sequence for chromosome III of *S. cerevisiae* became available for comparison with the conventional genetic map. Discrepancies between the DNA sequence and the recombination map included two genes that were placed in the wrong order by recombination analysis. In most organisms, therefore, a genomic sequencing program will require that the conventional genetic map based on recombination frequencies is supplemented by physical mapping procedures. These procedures include restriction mapping, which establishes the positions of restriction endonuclease recognition sites in a DNA molecule, locating markers on chromosomes by hybridizing marker probes to intact chromosomes, and mapping known sequences in genome fragments using PCR and hybridization.

8.2 Physical Maps

The molecular equivalents of linkage maps are physical maps that result from direct molecular analysis of genomic DNA. Toward the end of Chapter 7, we described how electrophoretic karyotyping allows sequences to be used as hybridization probes and be assigned to gel bands representing individual chromosomes as well as how those chromosomal bands can be removed from the gel to provide chromosome-specific DNA. This can be used for physical mapping of loci in a known, small region of the genome. The link between functional and physical mapping is the technique of restriction mapping. In this technique DNA fragments are digested with two or more restriction enzymes, and the fragments produced in this way are then compared by electrophoresis to establish a relatively short-range physical map. Restriction maps can then be the foundation for more extensive and more detailed maps.

There are a large number of restriction endonuclease enzymes, each with a different recognition sequence at which they cut DNA molecules. A restriction map can be constructed by comparing the sizes of the fragments produced when a DNA molecule is digested with two different restriction endonucleases. The basic principle of the technique is to digest the DNA molecule first with one of the enzymes, and then to measure the sizes of resulting fragments using agarose gel electrophoresis. The second step is to digest the original DNA molecule with the second enzyme and, again, measure fragment sizes by gel electrophoresis. Finally, the DNA molecule is digested with both enzymes at the same time, and the sizes of fragments produced by this double restriction determined.

These three sets of fragment sizes ideally enable the numbers and positions of the restriction enzyme recognition sites to be located in the DNA

molecule. Any inconsistencies can usually be worked out by using partial restrictions in which the digestion is prevented from going to completion by reducing the time or reducing the temperature for the incubation. The products of partial restriction include all the intermediate fragments that have one or more restriction sites still uncut; knowing the sizes of partially restricted fragments can help in putting the fully digested fragments together into a consistent map. Another additional source of information is to label the ends of the starting DNA molecule with radioactive or fluorescent markers before carrying out the restriction digestions. This allows the end fragments to be identified on the agarose gel.

8.3 Restriction Mapping: A Real-Life Example

As an example of restriction mapping, we will describe how the organization of the ribosomal DNA repeating unit from *Saccharomyces cerevisiae* was analyzed. This example shows both how restriction mapping works in practice as well as how specific functions can be associated with regions of the restriction map. It also serves to introduce rDNA, to which we'll return in discussing phylogenetics in Chapter 9. Let us start by looking at what was known when the study began. It was known that in yeast, as in other eukaryotes, the rRNA genes are repeated, with about 140 copies per haploid yeast nucleus, and at least 70% of these being located on chromosome I. The evidence indicated that there are equal numbers of genes for the 5S, 5.8S, 18S, and 25S ribosomal RNA subunits.

In yeast, the 5S rRNA genes are interspersed with the other rRNA genes. The RNA is transcribed from opposite DNA strands: 5S rRNA as a primary transcript, with the 5.8S, 18S, and 25S rRNA being processed from a 35S rRNA primary transcript precursor. The model of the yeast rDNA genes that existed when the restriction mapping was started was that 5S and 35S rRNA genes exist in blocks of tandemly repeating units, each unit of which contains one sequence for the 35S rRNA transcript and one sequence for the 5S RNA transcript. The 5S rRNA genes are linked with those for the large rRNAs in fungi as well as in a few other lower eukaryotes (e.g., like the cellular slime-mold, *Dictyostelium*), but they are unlinked in others (including the ciliate, *Tetrahymena*) and in higher eukaryotes (like the fruit fly, *Drosophila*, and the toad, *Xenopus*), even though the significance of this linkage of the 5S and large rRNA genes is not understood.

The restriction mapping we will describe was undertaken to establish the order, location, and polarity of the various rRNA coding sequences within the repeating unit of *S. cerevisiae*. The first task is to guarantee an adequate supply of the DNA fragment that is to be mapped, and this usually means cloning it. In this case, yeast DNA and an ampicillin-resistant plasmid were digested with the restriction endonuclease *Xma* I, joined together, and the construct was used to transform the bacterium *Escherichia coli*. Recombi-

nant clones were selected because of their ampicillin resistance, and those containing yeast rDNA insertions were identified by hybridization to yeast rRNA.

Ribosomal RNA genes attract a great deal of attention both because they are involved in a crucial aspect of cell metabolism and because the genes are repeated and their RNA products are present in enormous quantities in the cytoplasm. As a consequence of this repetition they are easier to isolate than other genes that might be present in the genome in single copy form and whose expression may generate only a few messenger RNA molecules. Purification of *S. cerevisiae* rRNA, therefore, is relatively straightforward. The 18S and 25S rRNAs were prepared from yeast cells from which most of the wall had been removed enzymically, then purified by sucrose density gradient centrifugation. Low molecular weight (5S and 5.8S) rRNAs were extracted from whole yeast cells and were separated by gel chromatography. Bearing in mind that these RNA molecules are to be used to identify the DNA sequences that will be the subject of all further analysis, their purity is a major consideration. This is always true, of course, in studies that rely on the potentially exquisite sensitivity of nucleotide base pairing; the probe that is being used to identify the molecule of interest must be demonstrated by independent tests to be sufficiently pure to justify the trust placed in it.

Here, the 5S rRNA was homogeneous (i.e., uncontaminated), whereas the cross-contamination of the other purified RNAs was less than 5%, which was judged to be adequate for this study. It was found that the restriction enzyme *Xma* I produced a fragment with a molecular weight of approximately 5.8×10^6 (= about 9300 base pairs) from yeast genomic DNA, which is the molecular weight predicted for the rDNA repeating unit. Hybridization analysis with the RNAs indicated that this DNA fragment encoded the 5S, 5.8S, 18S, and 25S rRNAs, which strongly suggests that this *Xma* I-generated fragment was the rDNA repeating unit. This DNA fragment was cloned, therefore, in a bacterial plasmid to provide the quantity of material needed for further analysis.

With supplies of the complete rDNA repeating unit assured, we can start attempting to digest it with restriction endonucleases. We say "attempt" because it's unlikely that the fragment in which we have an interest will contain a recognition sequence for every enzyme we're likely to try. For example, this DNA fragment was not digested by the restriction endonucleases *Sal* I, *Pst* I, or *Bam* HI, but the enzymes *Kpn* I, *Xba* I, *Bgl* II, *Hin*dIII, and *Eco*RI digested it. The fragments generated by these enzymes from the yeast rDNA repeat are summarized in Table 8.1.

The information we can extract from these data is limited to defining the number of recognition sites for each enzyme. One *Kpn* I site is evidently sufficient to generate two fragments; two *Xba* I sites will produce three fragments; three *Bgl* II sites are indicated, and seven *Eco*RI cleavage-sites. We can't map any of those sites yet because we do not have enough infor-

Table 8.1. Restriction fragments produced by digestion of the *Xma* I-generated rDNA repeating unit of *Saccharomyces cerevisiae*.

Restriction enzyme	Recognition sequence	Fragments produced (size in bp)
Original *Xma* I-generated fragment	5′-C↓CCGGG 3′-GGGCC↓C	9300
Kpn I	5′-GGTACC 3′-C↓CATGG	5600 and 3700
Xba I	5′-T↓CTAGA 3′-AGATC↓T	4500, 3800, and 1100
Bgl II	5′-A↓GATCT 3′-TCTAG↓A	4300, 3400, 1300, and 300
*Eco*RI	5′-G↓AATTC 3′-CTTAA↓G	2900, 2300, 2120, 600, 580, 300, 240*, and 240*

* These two *Eco*RI fragments can be separated from each other by electrophoresis in 5% acrylamide gel. ↓ = cleavage site in recognition sequence.

mation to arrange the fragments into their correct order. We need additional evidence to distinguish between the various possible arrangements, and, because polarity is important, we also need to know which of these fragments carries the cleaved *Xma* I recognition sites, which generated the original rDNA repeat unit.

So that we can identify the ends of the fragment in further digestions, the *Xma* I-generated DNA fragment was labeled with ^{32}P by adding [γ-^{32}P]-ATP to its ends using polynucleotide kinase. The labeled DNA was then digested with *Kpn* I, the two fragments produced separated by preparative electrophoresis, and the DNA recovered from the gels in sufficient quantity to supply the rest of the experiments. We now have two separate (*Kpn* I-generated) fragments of 3700 and 5600 bp, respectively, which we can subject to digestion with a second restriction enzyme. Each of these fragments is labeled at one end (i.e., the end with the *Xma* I recognition site) with a radioactive marker we can detect using autoradiography of agarose gels. These fragments were then subjected to partial digestion with *Eco*RI. A partial digestion with this enzyme is needed because it produces the largest number of fragments when used singly. The labeled *Kpn* I-fragments were also digested to completion with each of the other restriction enzymes to create double digestions.

For complete digestions, the restriction enzyme reactions proceed for 3 hours, so to make partial digestions the two *Kpn* I fragments were incubated with *Eco*RI for different times, but only up to 10 minutes. The radioactive partial digestion products were then analyzed by electrophoresis in agarose gel (Fig. 8.1). Remember that we are looking only at radioactive fragments, and that no fragment with an *Eco*RI cleavage site at each end will be detected because it will not be radioactive. Thus, all the detected fragments have an original *Xma* I-generated (and ^{32}P-labeled) end intact. As a result, the positions of the *Eco*RI sites from the radioactive end of the

fragment can be inferred from the sizes of the radioactively labeled partial digestion products.

The *Eco*RI partial digest of the 3700 bp *Kpn* I fragment produced four bands of approximately 240 bp, 480 bp, 2600 bp, and 3200 bp, corresponding to four *Eco*RI sites in this fragment (remember, the unlabeled fragments are not detected), located about 240 bp, 480 bp, 2600 bp, and 3200 bp from the ^{32}P-labeled *Xma* I site at its end. The 5600 bp *Kpn* I fragment similarly produced three radioactively labeled bands, corresponding to there being three *Eco*RI sites in this fragment that are approximately 2300 bp, 2900 bp, and 3200 bp from the ^{32}P-labeled *Xma* I site (which is, of course, at the opposite end of the rDNA repeat unit to the ^{32}P-labeled *Xma* I site of the 3700 bp *Kpn* I fragment).

Finally, a series of double digests were made by reacting the two radioactively labeled *Kpn* I-derived rDNA fragments with the other restriction

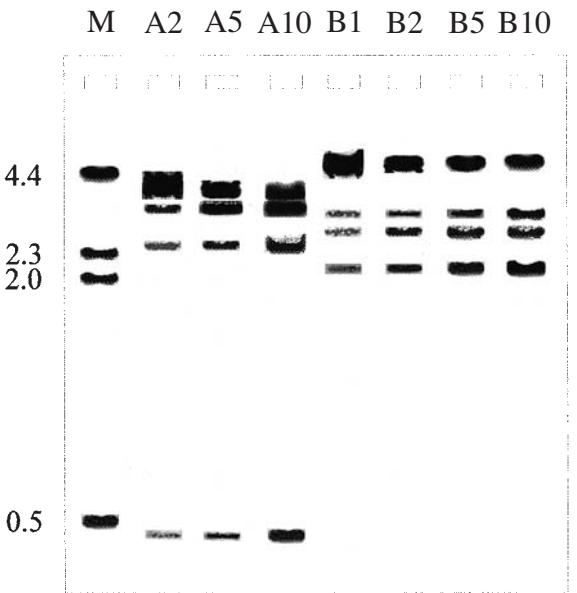


Fig. 8.1. Representation of an autoradiograph of partial *Eco*RI digestions of purified *Kpn* I-digested yeast rDNA fragments, radioactively labeled at the *Xma* I sites. The two *Kpn* I fragments were incubated with *Eco*RI for different times up to 10 minutes, and then the partial digestion products were separated by electrophoresis in agarose gel. Lanes labeled A2, A5, and A10 were loaded with DNA from the 3700 bp *Kpn* I fragment, and incubated with *Eco*RI for 2, 5, and 10 minutes, respectively. Lanes labeled B were loaded with DNA from the 5600 bp *Kpn* I fragment, and incubated with *Eco*RI for 1, 2, 5, and 10 minutes. Lane labeled M was loaded with the molecular weight marker DNA, which was *Hin*dIII-digested λ bacteriophage DNA. (Based on Fig. 3 in Bell, DeGennaro, Gelfand, Bishop, Valenzuela & Rutter (1977) *Journal of Biological Chemistry* **252**, 8118–8125.)

Table 8.2. Restriction fragments produced by double-digestion of the *Kpn* I-generated fragments of the rDNA repeating unit of *Saccharomyces cerevisiae*.

Second restriction enzyme	Fragments produced (size in bp)
Kpn I left-hand fragment	
Undigested	3700
*Eco*RI	2120, 600, 500, 240, 240*
Bgl II	2100, 1300, 300*
Xba I	2600, 1100*
Kpn I right-hand fragment	
Undigested	5600
*Eco*RI	2400, 2300*, 600, 300
Bgl II	4300*, 1300
Xba I	4500*, 1100

* Radioactively labeled terminal fragment.

enzymes. These reactions were all allowed to go to completion (i.e., the digestions lasted 3 hours) and all fragments, both radioactive and non-radioactive, were assayed. The approximate sizes of the fragments obtained are summarized in Table 8.2. This gives us enough information to complete the restriction map, and we show its development in Fig. 8.2. The two *Kpn* I fragments are key to our interpretations because their radioactive labels identify the extreme ends of the rDNA repeat. We need to orient these two fragments, so let's make the decision to place the smaller *Kpn* I fragment on the left and the larger on the right. This is just an arbitrary convention that enables us to anchor the two ends of our map on the page.

After we have drawn out the starting fragment and the two *Kpn* I fragments (Fig. 8.2), we use the partial digestion results coupled with the full double-digestion fragment sizes (Table 8.2) to place the *Eco*RI-derived fragments within the ribosomal unit into their correct order. The partial *Eco*RI-digestion of the 3700 bp *Kpn* I fragment gave two small radioactive fragments of 240 bp and 480 bp. These correspond to two *Eco*RI-cleavage sites, one of which is 240 bp from the left end of the fragment, and the other of which is 240 bp further away. These two sites yield the two 240 bp fragments found in the full digest (Table 8.2). The third *Eco*RI-cleavage site is approximately 2600 bp from the left end. This will be (2600 – 480) = 2120 bp from the second site, and the 2120 bp fragment in the complete digestion (Table 8.2) is produced by cleavage of these two sites.

The fourth, and last, *Eco*RI-cleavage site in the left-hand *Kpn* I fragment is 3200 bp from the left end, which means that it is (3200 – 2600) = 600 bp from cleavage site three. Cleavage at the third and fourth *Eco*RI-cleavage sites at the same time obviously yields the 600 bp fragment that is found in the complete digest (Table 8.2). When the fourth site is cleaved it produces

an end fragment that is (3700 – 3200) = 500 bp. A quick clerical check shows that we've been able to map all four cleavage sites and account for all the fragments found in the double digest.

Mapping the partial *Eco*RI-digestion of the 5600 bp *Kpn* I fragment proceeds in exactly the same way *except* that we now have to work from the extreme right-hand end because the partial digest catalogues fragments that were radioactively labeled at the right-hand *Xma* I cleavage site. The smallest radioactively labeled band produced in this partial digestion corresponds to an *Eco*RI cleavage site (which we are going to call site seven, so we can number all the sites in one sequence) that is 2300 bp from the right-hand end of the 5600 bp *Kpn* I fragment. Cleavage at this site generates the 2300 bp (and terminally labeled) fragment in the complete digestion (Table 8.2).

The next *Eco*RI cleavage site (site six) is 2900 bp from the right-hand end, which makes it 600 bp from site seven. Finally, cleavage site five is

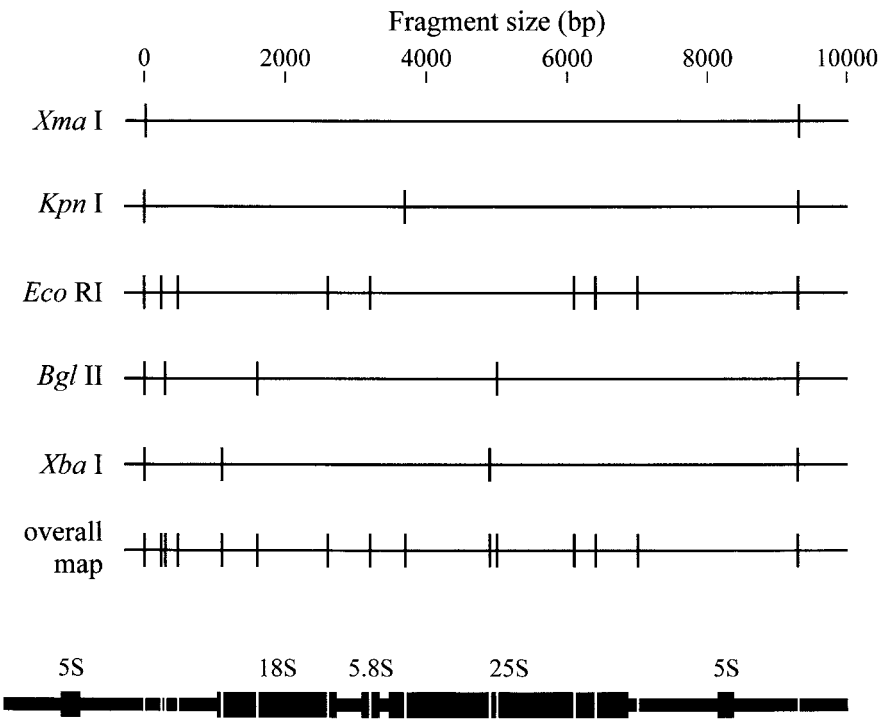


Fig. 8.2. Development of the restriction map of the yeast rDNA repeat. Positions of each restriction enzyme cleavage site, as deduced from the sizes of the fragments obtained (summarized in Table 8.2) are shown separately, and then the data are combined together to show the overall map. The bottom diagram combines the restriction map with an indication of the disposition of the genes corresponding to the different rRNA species, based on the hybridizations described in the text.

3200 bp from the right-hand end; that's 300 bp from site six and 2400 bp from the *Kpn* I site. Again, all *Eco*RI cleavage sites have been located and all the fragments found in the double-digest accounted for (Fig. 8.2). We have also accounted for all of the fragments generated in the *Eco*RI single digestion (Table 8.1), remembering that the *Kpn* I cleavage site lies between two *Eco*RI, so it is within an *Eco*RI fragment made up of the two end fragments from the partial digestion, which is an *Eco*RI fragment that is (500 + 2400) = 2900 bp in size.

Cleavage sites of the other enzymes can be mapped by combining the data shown in Tables 8.1 and 8.2. Thus, the first of three *Bgl* II sites is 300 bp from the left-hand end because the fragment of that size in the double digestion is radioactively labeled with the ^{32}P that identifies the left-hand *Xma* I site. For the same reason, the third (i.e., most rightward) *Bgl* II cleavage site is 4300 bp from the right-hand end, which is the radioactive fragment in the double digest (Table 8.2). The second *Bgl* II cleavage site could be either 1300 bp or 2100 bp to the right of the first site. The issue is decided by the single digest data in Table 8.1, which shows 1300 bp, and 3400 bp fragments that can both be generated if we locate the second *Bgl* II cleavage site 1300 bp to the right of the first. The positions of the two *Xba* I cleavage sites can be directly deduced from the sizes of the ^{32}P-labeled terminal fragments obtained by digestion of the two *Kpn* I fragments (Table 8.2).

Before we leave this research, we can add some functional data to the restriction map using hybridization between some of the DNA and RNA sequences we have in hand. We have the gels on which the various restriction enzyme digests were separated, and we have the purified ribosomal RNAs that were used to identify the genomic clones right at the start of the work. We can therefore now use the RNAs as hybridization probes to localize 5 S, 5.8 S, 18 S, and 25 S rRNA coding regions within the rDNA repeating unit. This analysis found that 18 S rRNA hybridized to the 2120 bp *Eco*RI-digested rDNA fragment, and 25 S rRNA hybridized to the 2900 bp fragment. This places the sequence of the 18 S RNA on the left-hand side of our restriction map.

5.8 S rRNA hybridized predominantly with the 2900 bp and 600 bp *Eco*RI-fragments, which places the 5.8 S sequence between the 18 S and 25 S sequences. The 5 S rRNA hybridized to the end fragment defined by *Eco*RI cleavage site seven and the *Xma* I site at the right-hand end of the repeat unit. DNA sequence analysis confirmed the RNA gene sequence and established the polarity of the 35 S rRNA transcript to be 5′-18 S, 5.8 S, 25 S-3′. Other data shows that 5 S and 35 S rRNAs are synthesized from DNA strands with opposite polarity (i.e., from complementary strands). The bottom panel of Fig. 8.2 summarizes the results of this restriction mapping of the yeast rDNA repeat and shows the yeast rDNA repeating unit to consist of a single gene for 5 S rRNA and the 35 S rRNA precursor, with the units being joined in a "head-to-tail" arrangement in the tandemly repeated array.

It must be evident from this example that restriction maps are easier to construct if there are only a few cleavage sites for the restriction enzymes being used. The more cleavage sites there are, the greater the number of fragments and the more complex the problem of establishing their order in the map. Interpretation can be aided with computer analysis, but large numbers of fragments increase the chance that several will have such similar sizes that bands merge on the agarose gel, making them difficult to measure correctly and making it impossible to assemble an unambiguous map. With a selection of restriction enzymes it is usually possible to construct unambiguous restriction maps of DNA molecules less than about 50 kb in length. Cloned sequences are often less than this size because of the capacity limitations of cloning vectors, so a detailed restriction map of a cloned fragment can usually be assembled, often as a preliminary to full sequencing of the cloned region.

These size limitations can be avoided to some extent by using "rare cutter" restriction enzymes. These are enzymes with recognition site sequences that consist of seven or eight nucleotides. Assuming random base composition, an enzyme with a seven-nucleotide recognition site is expected to cut once every 16,384 bp, and an eight-nucleotide enzyme once every 65,536 bp. Only a few such enzymes are known, but rare cutters can be used in restriction mapping of large molecules. A consequence of this technique is that the large DNA fragments that are produced may have to be separated using electrophoresis techniques similar to the PFGE technique we described for separation of whole chromosomes electrophoretically. Acronyms of related electrophoretic techniques you may encounter include orthogonal field alternation gel electrophoresis (OFAGE), contour clamped homogeneous electric fields (CHEF), and field inversion gel electrophoresis (FIGE).

8.4 Optical Mapping

Restriction maps are important because they can have an important role in the final assembly and verification of sequence data. The pattern of restriction enzyme cleavage sites described by a restriction map provides a series of landmark sequences at defined intervals, which enable fragments of the sequence to be assembled into a complete and finished sequence. Although sequencing costs have decreased because of automation, the cost of conventional restriction mapping remains high because of the enormous effort required to order restriction fragments, along with the facts that it requires skilled people and is time consuming and difficult to automate. High-resolution restriction maps can fortunately be made from single molecules by optical mapping without the need for electrophoresis. Optical mapping techniques quickly provide a high-resolution reference chart for aligning sequences accurately.

Optical mapping was first applied to prepare restriction maps of yeast chromosomes, and although the technique has been considerably adapted since, this analysis illustrates the underlying principle. For this approach chromosomal DNA in molten agarose was placed on a microscope slide that had been coated with a restriction enzyme, which was kept inactive because of a lack of the magnesium ions that are essential for restriction enzyme function. The DNA molecules become stretched out as the gel solidifies. The solidified gel was then washed with $MgCl_2$ to activate the restriction enzyme, and an intercalating dye that fluoresces when it binds to DNA. The fluorescence was sufficient for the DNA fibers to be visible with a conventional fluorescence (light) microscope. The technique works because the restriction enzyme digests the stretched molecules, after which the cleaved fragments relax and the cleaved restriction sites become physical gaps as the DNA fragments pull back on either side. The cuts, therefore, become visible as nonfluorescent spaces, but the fragments of the endonuclease-digested DNA molecules remain in position and the relative positions of the restriction cleavage sites can be measured and recorded. A 3 Mb DNA molecule will be 1 mm in length, and fragments as small as 800 bp can be imaged and sized.

Assembling the restriction map from such measurements made with a collection of restriction enzymes is essentially the same as we described earlier for conventional restriction mapping of the yeast rDNA repeat unit. The revolutionary difference is that the optical mapping system can be automated for high-throughput analysis, the observations can be made using computer-based image analysis, and sufficient data can be gathered for critical statistical analysis. Current preparation procedures use laboratory robots to place very small drops of DNA in defined 10×10 arrays that can be digested and analyzed together. The DNA molecules are oriented on the glass by circulating fluid flows developed within the tiny, evaporating droplets, which elongate and fix DNA molecules onto chemically modified ("silanized") glass surfaces, charge interactions holding the stretched DNA molecules on the surface. Machine vision and automatic image acquisition techniques have been developed to work with the gridded samples, and image analysis software produces high-resolution restriction maps from statistically meaningful populations of images of individual DNA molecules.

One of the surprising features is that optical mapping does create restriction maps from digital images of *single* DNA molecules; the final map is an average of many individual maps. The accuracy of the maps of individual molecules is affected by sizing errors, incomplete digestion, and false cuts (i.e., the optical system "seeing" gaps that are not really there) because of the limitations in resolution of the microscope, the imaging system, the surface conditions, and the efficiency of the restriction–digestion reaction (i.e., the percentage of available restriction cleavage sites that are cut by the enzyme), which can vary in the 70–90% range. The finished map is com-

puted using statistical approaches to find a maximum likelihood map. The operator can choose whether to combine data from 10, 100, or several thousand observations. One review article recommends that consensus maps be created using up to 8000 observations from the same data set. The latest software does this automatically, and will also accept the underlying sequence data, enabling real-time browsing of very large reference maps.

8.5 DNA Cloning: Plasmids, Cosmids, BACs, and YACs

Cloning involves inserting DNA molecules of interest into specialized carriers called *vectors* that enable replication within a host cell, producing many copies of the vector along with the inserted piece of DNA it carries. Cloning vectors are "engineered" to contain one or several recognition sites for restriction enzymes. Digesting both the vector and the DNA to be cloned with the same restriction enzyme produces complementary "sticky ends" in both molecules, allowing the foreign (or "heterologous") DNA fragment to be inserted into the vector. A vector carrying an inserted fragment of DNA is known as a *recombinant plasmid*. The replicated molecules are called *clones* because all the copies made in the host cell are identical. After harvesting from the host cell, the cloned DNA can be purified for further analysis.

There are several types of cloning vectors, which differ in origin, nature of host cell, and in their capacity for the size of inserted DNA they can carry. The simplest vectors are bacterial plasmids, which are circular, double-stranded, DNA molecules that replicate in the host independently of the main bacterial chromosome. Commonly used plasmids can carry up to 15 kb of foreign DNA. DNA fragments up to 25 kb in length can be accommodated in vectors derived from the bacteriophage ("phage") lambda (λ), which is a double-stranded DNA virus that infects the bacterium *Escherichia coli*. The λ phage replicates in its host as a circular molecule, but has a linear DNA molecule 50 kb long in mature (infective) virus particles. The virus chromosome circularizes after infection because it has complementary single-stranded overlaps at each end known as *cos* (for cohesive end) sites. A completely artificial, larger-capacity vector has been engineered by inserting *cos* sites into a plasmid. These are called *cosmids*. They can carry up to 45 kb of inserted DNA and have the additional advantages that they use a virus coat to infect host bacteria (a very efficient way of entering the host), but replicate like a plasmid and can be constructed to use plasmid-derived markers for recombinant selection.

A bacterial cloning system for analysis of much larger inserts has been developed from *E. coli* and its single-copy plasmid *F* factor, which carries inserted DNA in the form of bacterial artificial chromosomes (BACs). This system can clone DNA inserts of more than 300 kb, with a high degree of structural stability in the host, even after 100 generations of serial growth.

Because of the high cloning efficiency, stability, and easy manipulation of the BAC system, it has become the method of choice for cloning human libraries. Most large-scale human genome sequencing has used BACs carrying inserts of 100–200 kb as the main substrate for sequencing efforts.

The largest capacity vectors currently available are yeast artificial chromosomes (YACs), which can carry DNA inserts up to 1 million base pairs (1 megabase; 1 Mb) in length. YAC vectors are plasmids that contain yeast centromere DNA, two yeast telomeres separated by a restriction site, and yeast replication origins (autonomous replication sequences, or ARSs) as well as two selectable markers. Restriction enzyme digestion produces two fragments—one a telomere + selectable marker + cloning site, the other a telomere + selectable marker + replication origin + centromere + cloning site—that are mixed with the DNA to be cloned.

Among the constructs that result will be some that behave like yeast chromosomes during mitosis. Any that are constructed with two centromeres, without a centromere, or lacking a telomere will fail to segregate. As a result, the presence of both selectable markers coupled with proper mitotic segregation is sufficient to identify the desired constructs. It is interesting that at meiosis a diploid cell that contains two homologous YACs will show 1:1 segregations into progeny spores just like a regular pair of homologous chromosomes. YACs currently offer the highest capacity of any cloning vector, but recombination can affect insert stability in some. Choice of cloning vector, and consequently insert size, depends upon the purpose in mind when the cloning is initiated, as well as considerations about stability, reliability, and practicality. Storage of the library is one consideration because this depends on the vector. Plasmid, BAC, and YAC libraries are held in live cell cultures, whereas phage and cosmid libraries are suspensions of mature virus particles refrigerated in a form ready to infect their host.

The capacity of cloning vectors is important for two reasons. First, the genetically functional regions of some higher eukaryotes are large, often because the genes contain many introns. Thus, if it is desired to clone the genomic DNA, a large-capacity vector is required. Second, the larger the capacity of the vector, the fewer the total number of clones that are needed for a library that is representative of the entire genome. Clone libraries are random collections of the DNA fragments you obtain by restriction digestion of the genome, subsequently inserted into a suitable vector and cloned. The number of clones the library must contain to be equivalent to one genome can be calculated by dividing the size of the genome by the average size of the inserts carried by the vector you want to use. To clone a representative fungal genome of 30 Mb in a phage vector with a 25 kb capacity would require a minimal library of 1200 clones. Of course, if you aim to generate only 1200 clones it is impracticable to expect that every DNA fragment will be represented once, and once only, in a random sample of DNA from a restriction digest. You would expect that some DNA frag-

ments would appear more than once, and that others might not be present at all.

To improve the probability that a library is fully representative, the number of clones in the library must exceed this minimal value. Aiming for four to five genome equivalents gives a 95% probability that any individual locus will be present at least once; therefore, a phage library would have to contain 6000 clones to meet this requirement. At the other end of the spectrum, if you could reliably prepare genomic fragments averaging 1 Mb, you would only need a clone library of 30 YACs for the minimal library; or 120 for a 95% probability that every locus is represented. For a real example, consider the 12.8 Mb yeast genome, and assume an average insert length of 35 kb in a cosmid. A cosmid library containing 4600 random clones would have a probability of 99.99% of representing the entire yeast genome. But $4600 \times 35\,\text{kb} = 161\,\text{Mb}$, which is just more than 12 times the genome equivalent; thus, a lot of clones are required if you want to be absolutely certain that every sequence is represented in your library.

Although large inserts mean that fewer clones have to be analyzed to find a DNA sequence of interest, very large inserts are often difficult to handle in standard procedures. Subcloning is used to avoid this disadvantage. The large insert that includes the DNA sequence of particular interest is digested with restriction enzymes and the smaller pieces cloned into a different vector. Successive rounds of subcloning can establish sublibraries of fewer and fewer clones that are enriched for the sequence of interest.

8.6 Chain Termination Sequencing

The ultimate physical map of DNA is its nucleotide sequence. There are two basic procedures for DNA sequencing, both of which were first published in 1977: (1) the chain termination method (or dideoxy-sequencing) sequences a single-stranded DNA molecule using enzymic synthesis of complementary polynucleotides that terminate at particular nucleotide positions when a nucleotide triphosphate lacking a 3′-hydroxyl group is incorporated; (2) the chemical degradation method uses different chemicals to cleave the DNA after a particular base or bases (either A, G, T, or C). Both procedures generate a complete set of sequence fragments that differ in length by just one nucleotide from both the preceding and succeeding fragments, a collection of fragments known as a *nested array*. This array of DNA molecules differing in length by one nucleotide, and potentially covering a size range from 10 to 1500 nucleotides, can be separated by polyacrylamide gel electrophoresis. The end nucleotide is identified by some sort of label added to it during the sequencing procedure. The chain termination procedure has become the principal technique because it has been possible to automate the process to the extent that large-scale sequencing can be completed at reasonable cost and in a reasonable time-

scale. Chain termination may have surpassed the chemical degradation method as the method of choice for automation because the latter suffers from problems in developing chemical reactions that cut at A and T bases specifically.

Chain termination sequencing relies on the synthesis of new strands of DNA complementary to a single-stranded template, which is the target DNA. As a result, a clone of the DNA molecule to be sequenced, prepared in single-stranded form to serve as the DNA synthesis template, is the starting material for chain termination sequencing. The clone may be made as part of the subcloning procedure from a BAC or YAC library. The most common method is probably to use cloning in a plasmid vector. This yields double-stranded DNA, which is denatured to single strands with alkali or by boiling. Both complementary single strands produced can then be analyzed to provide independent sequences of both ends of the cloned DNA. The DNA can alternatively be cloned in a vector based on the M13 bacteriophage, which is designed specifically to produce single-stranded templates for DNA sequencing.

Mature M13 bacteriophage particles contain the single-stranded copies of the cloned DNA molecules. Purification is an issue with plasmid vectors because the plasmids have to be harvested from whole cells and then rigorously purified from other polynucleotides. M13 bacteriophages are more easily purified; however, insert size becomes an issue because inserts longer than about 3 kb can suffer deletions and rearrangements when cloned in M13. A compromise vector is a phagemid, which is a plasmid containing an M13 origin of replication. When combined with a helper phage, single-stranded copies of the phagemid cloning construct are packaged into virus particles. Phagemid vectors can accommodate inserts up to 10 kb. PCR can also be used to prepare template DNA and one of the PCR primers can then be used as a primer for the template-dependent DNA polymerase.

Once you have your template, the first step in sequencing is to anneal a short oligonucleotide to it, which will serve as the primer for DNA synthesis by a DNA polymerase enzyme. The sequence of this primer determines where the sequencing process will start, so it is in the experimenter's hands to decide which primer is appropriate to the sequence of interest. In addition to the enzyme itself, DNA synthesis requires the four deoxyribonucleotide triphosphates (dATP, dCTP, dGTP, and dTTP) as its usual substrates. The reaction mixture also contains small amounts of a dideoxynucleotide, which has a hydrogen atom rather than a hydroxyl group on its 3′-carbon atom. The DNA polymerase incorporates the dideoxynucleotide without trouble, but is then blocked from further elongating the growing chain. This is because the 3′-OH, which is normally phosphorylated and linked to a 5′-sugar carbon atom on one side and a 3′-sugar carbon atom on the other side to form a double-stranded polynucleotide polymer, is missing in the dideoxynucleotide. That is the "chain termination" part of chain termination DNA sequencing.

Let us look at the results of carrying out this DNA synthesis in the presence of dideoxy-ATP. Including ddATP in the reaction means that DNA daughter chains will be terminated at random at all those points where the template has a thymine. This will produce a family of adenine–nucleotide-terminated chains that have lengths equivalent to all the stretches in the template that extend from the 3′-end of the primer to each thymine in the sequence. Thus, between them these fragments report the position of every thymine in the template. If this family of fragments is electrophoresed in polyacrylamide gel, the fragments in the population will migrate at a rate dependent on their exact length, and a series of bands will be obtained where each band corresponds to one of those "3′-end to thymine" stretches. Of course, there are four nucleotide bases, so there will be four such reactions, and, alongside the "3′-end to thymine" family, there will be three other families of terminated chains corresponding to the "3′-end to adenine," "3′-end to cytosine," and "3′-end to guanine" stretches.

The banding pattern was originally visualized by autoradiography, achieved by including a radioactively labeled nucleotide in separate strand synthesis reactions. Products of the four reactions (i.e., the ddATP, ddCTP, ddGTP and ddTTP-terminated reactions) were loaded into adjacent lanes of the gel. The smallest molecules ran fastest during electrophoresis, so the sequence could then be read from the bottom of the autoradiograph by noting the position of the band in any one of the four lanes. Bands in the lane loaded with the ddATP reaction products reported the positions of thymine in the template, bands in the lane loaded with ddCTP-products reported the locations of guanine, ddGTP reported cytosine, and ddTTP-terminated reactions reported adenine locations in the template. The sequence could be read from its 3′-end by reading up the four lanes of the gel. The smallest, fastest running molecule represented an oligonucleotide terminated at the first base position after the primer site in the template, the second band corresponded to the second base position after the primer, and so on. The sequence of the DNA template could be deduced by continuing to read the banding pattern upward on a gel until the point at which the bands could no longer be resolved, near the top of the gel, and corresponding to about 1500 nucleotides in a good separation.

This procedure is technically undemanding, but as described earlier is time-consuming and labor-intensive. Over the years various aspects have been automated, but the real breakthrough came with the development of dideoxynucleotides labeled with four different fluorescent labels, which allow the dideoxy-terminated bands on the gel to be distinguished by the color of their fluorescence. The fluorolabels are attached to the dideoxynucleotides, a different fluorochrome is used for each one, and they avoid the more arduous and hazardous autoradiography. Fluorolabeling enables all four dideoxynucleotides to be used in one reaction tube, but it has also permitted the development of sequencing machines, which have a fluorescence detector that can discriminate between the different labels. Sequenc-

ing machines can carry out the sequencing reactions, use capillary separation rather than polyacrylamide gels, and screen the fluorescence of the separated bands as they emerge and pass in front of the detector, identifying and recording the sequence automatically. The fluorescence detection system offers a significant improvement in throughput and avoids the operator errors that can result when a sequence is read by eye and manually entered into a computer. Accuracy of the procedures is approximately 99.9% (= 1 erroneous base for every 1000 bases sequenced), and even greater accuracy is achieved if the complementary strand of the target DNA molecule is also sequenced.

Even with the automation of chain termination sequencing, the limitation in data yield from each run provides a spur to further development. This has led to the development of more radical approaches. Pyrosequencing uses a flash of chemiluminescence that accompanies release of pyrophosphate as each nucleotide is incorporated into the new strand. It enables the sequence to be read as the reaction proceeds and is more rapid than chain termination sequencing because it does not require fragment separation techniques like electrophoresis. There is also a drive to adapt the sequencing process to computer microchips. It seems likely that DNA arrays could be adapted to DNA sequencing, raising the possibility that microchips carrying arrays of different oligonucleotides could be developed to establish sequences by determining the hybridization patterns of the test molecules to the components of the array. Continued advances in miniaturization and computer data processing, combined with electronic detection of hybridization, could make such arrays a viable means of sequencing large molecules. An existing example of an imaginative new way of studying fungal genomes is an array comprising DNA from strains subjected to systematic directed mutagenesis. Developed by a company called Paradigm Genetics, the method is called Transposon Arrayed Gene Knock Out (TAG-KO™) and it is claimed that it can yield a "fungal phenome," which is a phenome that is the total of discernable phenotypes that arise from a genome wide mutagenesis program.

In spite of the development of automated techniques, on average at the moment, a single sequencing run can establish the sequence of a fragment from a few hundred to just less than a thousand bases long. Longer sequences have to be "stitched together" from smaller ones, and there are three basic strategies for this: shotgun sequencing, clone contigs, and directed sequencing.

Shotgun sequencing assembles the complete sequence directly from the output of sequencing experiments by looking for overlaps between the shorter fragments. It can be done without prior knowledge of a genetic or physical map, but it does require that many fragments are sequenced, and computer matching of the many short sequences represents a major data manipulation task. The target DNA sequence, which may be genomic DNA, or a BAC or YAC clone, is broken into random fragments with restriction

enzymes or by physical treatment like shearing or sonication. The fragments are then electrophoresed and those in the 2 kb size range cloned in a plasmid or phage vector.

You may then be faced with thousands of sequencing experiments, leading to five or six times the sequence length of the target sequence. This level of redundancy is necessary to ensure complete coverage; statistically, generating a cumulative sequence that is 6.5 to 8 times the true length of the molecule of interest guarantees more than 99.8% coverage of that sequence. Sequence assembly is done by computer analysis with programs that compare the short sequences, find overlapping ends, and put together the contiguous sequences (or sequence contigs). The shotgun strategy has been applied with great success to genome sequencing of several prokaryotes. It can be adapted to something like a production-line, with each team member having a specific (and highly repetitive) task to perform. It is so successful that the general expectation now is that 1 year is sufficient time to establish the complete sequence of any genome less than about 5 Mb!

A serious weakness of the approach is that every sequence that is determined must be compared with every other sequence to identify overlaps, so the data analysis needed to find the sequence contigs is complex. If you are interested in particular genes for which you have some sequence information (and, therefore, hybridization probes), it is possible to use sequence skimming in the early stages of a shotgun-sequencing project. This approach aims at concentrating attention on aspects of particular interest rather than sequencing an entire molecule, and relies on the expectation that there is a good chance that the first few sequence fragments, because they are obtained at random, will include recognizable parts of the genes of interest. It may be that this identification of the genes present in the cloned molecule satisfies the objective. Otherwise, a detailed sequencing effort can be focused on the genes of interest rather than on other parts of the cloned molecule. Several other short cuts to greater efficiency have been found, but more structured strategies are required for larger sequences.

The *clone contig* strategy uses information from previously established genetic and/or physical maps. This approach applies the shotgun sequencing strategy to libraries of overlapping cloned fragments derived from restriction digests directed by previously established maps. This is described as "anchoring" the fragments onto a genetic and/or physical map. The sequence data from the clones can consequently be ordered (made "contiguous") by reference to the pre-existing map; this is the "clone contig" aspect of the strategy. This approach is well suited to BAC and YAC clones.

Clone contigs are built up by using the DNA sequence from the starting clone to make a hybridization probe that is used to screen the rest of the library to identify a second clone with which the first overlaps and then the sequence of the second to identify a third clone with which the second overlaps, and so on. This is the essence of chromosome walking. The procedure

can be confused if the target DNA contains many repetitive sequences (relatively uncommon in fungi), which is a problem that can be avoided if short fragments from the ends of one fragment are used as probes to find overlapping clones. The slow process of chromosome walking can be avoided if high-resolution restriction maps are available, or are prepared, in sufficient detail to act as fingerprints of the clones. Clone contigs can then be assembled from clones that have restriction fingerprint features in common in their overlap regions.

The *directed shotgun approach* makes use of the sequence as it is obtained. Sequence information for the end of each sequence is used to synthesize PCR primers, which are used to recover the adjacent sections. The process can be repeated until the entire region of interest has been sequenced and assembled.

Of course, these different approaches are complementary, rather than mutually exclusive, and can be used in combination to fit the needs of the project. The ultimate goal of a geneticist is the complete description of the genetics of an organism. A genetic map is only one aspect of that "complete description of the genetics," but current molecular technology does make it feasible to undertake a genome project with the aim of determining the complete nucleotide sequence of every chromosome in the genome of *your organism*. During the mid-1980s the word *genomics* was coined to describe the "new science" of mapping, sequencing, and analyzing genomes. As we have described it, sequencing of microbial genomes may appear to be fairly routine, but it is a major undertaking. Approximately 600 scientists were involved in the yeast genome-sequencing project over a period of about 6 years. That is more than 3000 person-years of sequencing effort on an organism that had regularly featured in research over the previous hundred years and for which, at the outset of the sequencing project, we already had a conventional linkage map featuring some 1200 genes encoding either RNA or protein products.

Because of their generally smaller size, prokaryotic genomes were the first to be completed and work with these showed that comparison of genomes helps to answer some fundamental biological questions—and raise many others! Comparison of DNA and protein sequences has been made possible only by the development of computer programs that enable the sequence data to be stored and analyzed effectively. This is a new branch of biology called *bioinformatics*. The priority of genomics is to establish the number and function of genes in an organism. As a genome sequence is assembled the functional genes in the sequence are recognized as open reading frames (ORFs). Not all of the ORFs that are identified can be associated with a gene of identified function; an ORF specifying a product that does not resemble a known protein is called an unidentified reading frame (URF). Comparative genomics, however, does more than identify the genes. It can show the evolutionary relationships between different organisms, and

it aids the understanding of how the genotype relates to lifestyle and environment.

Physical and molecular analyses originally moved the genetical focus away from the functional gene and toward the DNA sequence; now, functional genomics have become the new buzzwords and an approach that is attempting to reverse that course, seeking instead to integrate biological understanding. Sequencing an entire genome is currently seen as the beginning of functional studies of the transcriptome (all the transcripts made from the genome), the proteome (all the polypeptides made from the transcriptome), and the metabolome (all the metabolic reactions governed by the proteome). Now, at the beginning of the twenty-first century, our understanding is that the genome is not context sensitive because it is the full set of genetic information. Instead, the transcriptome, proteome, and metabolome are all context sensitive because what they comprise depends upon the instantaneous regulatory status of the cell. As the old-time geneticists at the beginning of the twentieth century would put it: phenotype = genotype + environment!

8.7 The First Complete Eukaryotic Genome Sequence: *Saccharomyces cerevisiae*

The yeast genome-sequencing project was started in 1989. The sequence of chromosome III was the first to be published in 1992, chromosomes II and XI followed in 1994, and the sequence of the entire genome was released in April 1996.

Quality control measures ensured a 99.97% level of accuracy of the sequence. The yeast genome comprises 12.8 Mb, distributed over 16 chromosomes that range in size between 250 kb and more than 2.5 Mb. Approximately 70% of the complete genome sequence is taken up by 5800 ORFs, which possibly encode specific proteins. That means that, on average, a protein-encoding gene is found every 2 kb in the yeast genome. The genome sequence is still being actively analyzed and interpreted. As of May 2001, the reassessment of the *S. cerevisiae* genome revised the gene number estimate to an upper limit of 5570 (excluding very hypothetical genes and pseudogenes); for comparison, remember that the genomes of the bacteria *Escherichia coli* and *Bacillus subtilis* contain 4290 and 4107 protein-coding genes, respectively.

The value of 5570 for yeast is still predicted to be an overestimate of gene number. The ORFs vary from 100 to more than 4000 codons, although two-thirds are less than 500 codons, and they are fairly evenly distributed on the two strands of the DNA. In addition to these, the yeast genome contains 120 rRNA genes in a large tandem array on chromosome XII, 40 genes for small nuclear RNAs, 274 tRNA genes (belonging to 42 codon families) scat-

tered across the chromosomes, and 51 copies of the yeast retrotransposons (Ty elements). There are also nonchromosomal elements, most notably the yeast mitochondrial genome (80 kb) and the 6 kb 2 μ plasmid DNA, but there may be other plasmids, too.

Several surprising features have emerged from analysis of the architecture of the yeast genome, and as this analysis continues there are no doubt more surprises to come. *S. cerevisiae* genes have few introns. Fewer than 5% of the protein-encoding genes are interrupted by introns. There is usually a single intron (only two genes have two introns), which is generally at the extreme 5′-end of the gene, sometimes even before its coding region. This lack of introns is an exceptional feature of *S. cerevisiae*; the genes of other fungi, including all filamentous ascomycetes that have been studied, contain more introns. Even the genome of the fission yeast, *Schizosaccharomyces pombe*, has a lower gene density (one gene per 2.3 kb) than *Saccharomyces cerevisiae*, and about 40% of fission yeast genes contain introns.

With ORFs making up 70% of the *S. cerevisiae* genome, the high gene density can result in the intergenic regions between consecutive ORFs being extremely short. This leaves only limited space for regulatory sequences involved in DNA transcription, replication, and chromosome maintenance. A variety of transcription control elements have been identified, including upstream activation sequences (UAS) and upstream repressing sequences (URS). Some terminator sequences have also been defined, but no consensus sequences are evident. Regulatory elements are not confined to intergenic regions; they can also be located within the coding sequences of upstream neighboring genes. Such an arrangement clearly imposes evolutionary constraint on the sequence because of its dual function: selection must operate on the DNA sequence in relation to both the function of the protein specified by its coding sequence and its ability to regulate a gene of unrelated function some way downstream.

About 66% of the total ORFs represent novel yeast genes of unknown function; these currently undiscovered genes have been called *orphans*. Approximately 2300 ORFs (more than 40%) specify yeast membrane proteins, and although many of these fall into families (e.g., 33 mitochondrial transporters, 200 sugar and amino acid transporters, etc.), about 1600 of them are unique, with no homologues elsewhere in the genome. Given the importance of membrane processes in eukaryotes, devotion of such a large proportion of the genome to membrane proteins may not be unexpected.

A considerable surprise, though, is the level of genetic redundancy in yeast. Up to 40% of the gene sequences are duplicated. In most cases the duplicated sequences are so similar that their protein products are identical and, presumably, functionally redundant. These redundant proteins can substitute for each other if one is mutated, which explains why so many experimental single gene disruptions in yeast do not impair growth or cause abnormal phenotypes. A wide variety of these identical genes locate to different chromosomes. Examples include histone genes, genes for ribosomal

proteins, ATPases, amino acid and sugar transporters, and genes for enzymes of the glycolytic pathway. A sequence difference in the promoters of duplicated genes implies differences in expression, however, so expression of the different copies may depend on the nutritional or differentiation status of the yeast cell. None of these duplications appear to be pseudogenes, of which there are rather few in the yeast genome. Chromosome I, which is the smallest eukaryotic chromosome so far known, is exceptional in having four pseudogenes at each end. The pseudogenes are related to known yeast genes, but they contain internal stop codons.

More surprising than duplicated genes are numerous large segments on two or more chromosomes that share duplicated genes arranged in the same order and with the same transcription orientations. These are called cluster homology regions (CHRs) and there are 50 of them in the yeast genome. Ten of these CHRs (shared with chromosomes II, V, VIII, XII, and XIII) are located on chromosome IV, and the whole of chromosome XIV is made up of regions duplicated on other chromosomes. Outside the coding regions of these clusters the DNA sequence has diverged, which implies that the duplication events are ancient. The greatest level of duplication occurs in genes of unknown function. Duplication of metabolic proteins has not occurred on a major scale, but genes for proteins involved in membrane processes, control of protein conformation, and in DNA or RNA processing are highly redundant. This might mean that duplications improve environmental fitness by affecting integration and coordination of the major metabolic functions. It is interesting that genes that are crucial to the most basic cell functions (i.e., protein conformation, membrane transport and DNA/RNA processing) are also surmised to have arisen from ancient duplications, which suggests that there has been a definite drive to conserve these sequences throughout their evolution.

An important point to make is that the yeast genome sequence we have been discussing is that belonging to a specific laboratory strain, code number αS288C. As such it may have been affected by unconscious artificial selection during its "domestication," so it may or may not be representative of the natural populations of *Saccharomyces cerevisiae*. Some comparisons can be made from partial genome sequences. A significant fact is that sequence variation in the coding regions of individual genes is rare and does not contribute significantly to polymorphism between strains of *S. cerevisiae*. Rather, polymorphisms between yeast strains are particularly caused by differences in the number of gene copies within families of repeated genes, the distribution of Ty elements, and variation in the genetic redundancy and telomeric repeats that are found at all chromosome ends. Chromosome restructuring also differentiates yeast strains. Chromosome breakage can cause an altered karyotype, or deletions that give rise to chromosome-length polymorphisms.

Because of its established role as a model organism, particularly in cell biology in relation to cancer studies, there is an expectation that the avail-

ability of the complete yeast genome sequence will enable the identification of genes relevant to disease in humans. Comparing the sequences of human genes available in the sequence databases with yeast ORFs shows that more than 30% of yeast genes have homologues among the human sequences, with most of these representing basic cell functions.

Finding this sort of homology can contribute to the understanding of human disease. The example of Friedreich ataxia is interesting in this respect. It is the most common type of inherited ataxia (loss of control of bodily movements) in humans, but the biochemistry underlying it was uncovered by demonstrating homology to a yeast ORF of known function. Friedreich ataxia is caused by enlargement of a GAA repeat in an intron that results in decreased expression of the frataxin gene. *Frataxin* is the human mitochondrial protein that has homologues in yeast. In yeast, mutants defective in the frataxin homologue accumulate iron in mitochondria and show increased sensitivity to oxidative stress. This suggests that Friedreich ataxia is caused by mitochondrial dysfunction and may point toward novel methods of treatment. In many ways, this kind of comparison alone can justify all the effort devoted to sequencing the yeast genome. Without belittling this contribution in any way, our attention must concentrate on the fungi.

8.8 Comparisons Between Genomes

We can't yet compare complete genome sequences of fungi. At the time of writing, extensive sequencing projects are underway with several fungi, but none has matured to the point where large-scale comparisons are worthwhile. We can give an indication, however, of how the programs are progressing and what (and in some cases, when) outcomes are expected.

Once a sequence has been established and verified, the process of "annotating the genome" starts; that is, associating its component sequences with specific functions, and, if the *S. cerevisiae* example is a guide, continues for a long time. Annotation requires sophisticated computation. Gene identification is probably the most difficult problem, and it relies heavily on computer algorithms that align sequences and "*de novo* gene finder" programs. The gene finding problem has been largely solved for bacterial genomes in which the "Glimmer" system can find 97–99% of all genes automatically, but both gene finding and gene function assignment remain difficult tasks for eukaryotic sequences. Sequencing projects reveal "new" genes at a high rate. Determining the cellular role of these new open reading frames is the next major problem. Bioinformatics needs to integrate sequence information with the accumulated knowledge of metabolism so that conjectures can be made about likely functions of previously uncharacterized genes. Those predictions can then be tested experimentally, using heterologous expression, gene knockouts, and characterization of purified proteins. Parallel

analysis of phylogenetically diverse genomes can also help in understanding the physiology of the organisms.

The majority of genes identified in genome projects have an unknown function, yet, paradoxically, these genes are perceived to constitute a large pool of potentially novel targets for control and/or manipulation. This applies whether the organism concerned is a pathogen, for which "control" implies downregulation of growth, or a biotechnologically important crop fungus, for which "control" implies upregulation of growth. Major research efforts are being aimed at developing new computer strategies for efficiently analyzing and comparing completely sequenced genomes regardless of size or number of genomes. The goal is for these systems to identify essential genes and functional domains, and predict biochemical and cellular function. Applied to disease research, such systems could evaluate potential genetic targets for control strategies.

As we will see shortly, there is a considerable investment at the moment in practical aspects of genomic research, especially medical aspects, but including plant pathogens, too. Very large scale analysis of the very large data sets represented by genome sequences (the very essence of bioinformatics research), however, is likely to have an enormous impact on our understanding of the basic aspects of biology. An example of this is the analyses that reveal that eukaryotic cell nuclei originated by symbiosis of archaea within bacterial cells. A search for homology across the genomes of the eukaryotic yeast, six archaea and nine bacteria partitioned yeast gene homologies between either archaea or bacteria. Nucleus-related genes (e.g., the ones involved in transcription, replication, cell cycle, nuclear architecture, or ribosome biogenesis) shared closest homology with archaeal genes rather than with bacteria. In contrast, yeast genes that have cytoplasmic functions (e.g., metabolism, stress response, detoxification, or protein and ion transport) were more closely homologous to bacterial genes than they were to archaeal genes. This clear demarcation suggests that the eukaryotic nucleus is of archaeal origin. This leads to the conclusion that the eukaryotes arose from an endosymbiosis of an archaeal cell within a bacterium, in which the former lost its metabolic functions and the latter lost its own gene expression functions.

This sort of insight can only be obtained from extensive analysis of whole genome sequences, and it should not be too long before the same sort of broad comparison can be made within the fungi. Shotgun sequencing of the diploid *Candida albicans* has already provided a partial genome sequence, which has been used to make a search for homologues of genes known to be involved in meiosis in *Saccharomyces cerevisiae* and some higher eukaryotes. This study identified homologues of genes crucial for mating and meiosis in budding yeast within the genome of *C. albicans*. Taken together with the identification of sequences homologous with the *S. cerevisiae* mating type locus, this suggests these two fungi may have similar sexual cycles. *C. albicans*, however, is generally considered to be completely

asexual. If the molecular indications can be upheld, and the presence of a sexual cycle in this important human pathogen can be confirmed, then a significant hindrance to conventional genetic analysis in this organism will have been removed.

Saccharomyces cerevisiae is obviously the best-studied fungus, and the fission yeast *Schizosaccharomyces pombe* is also an important model organism for which a complete genome will soon be available. Neither of these yeasts, however, is an adequate model for filamentous fungi, which have more genes (approximately 8400) and bigger genomes (30–40 Mb); both features are presumably related to the wider morphogenetic, metabolic, and ecological capabilities of filamentous fungi. It is certainly already clear that several genes present in *A. nidulans* are not present in *S. cerevisiae*, so we await with keen anticipation the opportunity to compare genomes of filamentous ascomycetes. There are several genome projects underway; including the two filamentous ascomycetes that have been the most important genetic models, *Aspergillus nidulans* and *Neurospora crassa*. Among others, *Aspergillus fumigatus* is emerging as a major human pathogen, causing more infections worldwide than any other mold, and the rice blast fungus, *Magnaporthe grisea*, is a major plant pathogen.

The *Neurospora* Genome Project is also progressing. ESTs have been obtained for more than 2000 different genes that are expressed at different vegetative or sexual stages, or during different intervals of the circadian cycle. More than half of these have no known homologues in the yeast genome or elsewhere. Physical maps of the genome are under construction, and genetic mapping has progressed substantially using conventional markers, RFLPs, and chromosome rearrangements. The haploid genome of *N. crassa* contains approximately 43 Mb of chromosomal DNA; individual chromosomes range from 4 to 10.3 Mb. The seven linkage groups have been identified cytologically with individual chromosomes and assigned to separate DNA molecules using pulsed-field gel electrophoresis. There is little repetitive DNA in the *N. crassa* genome, mainly the genes specifying ribosomal RNA. *N. crassa* telomeres have a DNA sequence identical to that of humans. The Whitehead Institute's Center for Genome Research release schedule for the *Neurospora crassa* genome includes a conservative initial assembly of 1705 contigs containing 38,244,162 base pairs forming the first draft of the *Neurospora crassa* genome, which was released in February 2001. An improved assembly integrated with cosmid and BAC sequences and correlated with the genetic map was scheduled to be released in the summer of 2001, with a fully annotated genome due in December 2001 or January 2002.

The physical (contig) map describing the 31 Mb genome of *A. nidulans* has been completed, with cosmid contigs and EST mapping, and anchored to the genetic linkage map at the University of Oklahoma. In *A. nidulans*, as in *N. crassa*, preliminary data indicate that about half of the genes discovered are "new" in the sense that their sequences have not been encoun-

tered previously nor associated with a function in any organism. Very little is known about the biology and genetics of *Aspergillus fumigatus*, but 50 complete genes have been cloned from a genome estimated at 30 Mb. An incomplete electrophoretic karyotype shows at least five bands with two small chromosomes of 1.7 and 1.9 Mb, although some karyotype polymorphism is evident. A BAC library is in preparation and plans envisage anchoring clones to a physical map with ESTs. *A. fumigatus* is an extremely important human pathogen, causing allergic diseases in asthmatic and cystic fibrosis patients, as well as invasive aspergillosis in immunocompromised patients and those suffering from tuberculosis or other cystic lung diseases. A pilot genomic analysis of *A. fumigatus* is underway with an anticipated completion date of 2003.

Another important lung pathogen is *Pneumocystis carinii*, which causes pneumonia in immunocompromised patients and is consequently a major infection risk to patients with HIV infections and individuals undergoing organ transplantation, chemotherapy, or those with congenital deficiencies. *P. carinii* is not sensitive to the usual antifungal agents. It is hoped that genome analysis will lead to identification of more effective targets for new treatments. Much the same is expected of the *Candida albicans* genome project, which is also at a preliminary stage.

Magnaporthe grisea causes rice blast, and crop losses have been magnified in recent times as rice production has intensified; enough rice is lost to this disease to feed 60 million people each year. Strains of the fungus also attack other cereals, including wheat and barley, and it is a serious disease of turf grasses. Genetic resistance in the host plant has been and continues to be the major means of disease control for blast, but *M. grisea* is able to evolve rapidly and overcome major gene resistance. The goal of genome analysis of the fungus is to understand fungus–host interactions well enough to develop durable, and environmentally sound strategies to manage rice blast disease. A BAC library, which contains inserts that average 130 kb, has been prepared of *Magnaporthe grisea* to initiate contig assembly, gene cloning, and analysis of the seven chromosomes in the 40 Mb genome of this fungus. There are already several high-density molecular linkage maps, containing several hundred RFLP markers as well as several functional gene markers, which can provide a framework for anchoring contigs.

Plans have also been announced to sequence *Phytophthora* genes expressed during infection and propagation. *Phytophthora* species attack almost every dicotyledonous plant species, causing many billions of dollars damage to crops worldwide every year. The immediate goal is to focus on *P. sojae* and *P. infestans*, which have served as models for the genus in many studies over the years in the hope that will greatly advance the understanding of the biology and pathology of all *Phytophthora* species, providing tools for developing novel control measures against these crop pathogens. A similar project plan is aimed at the discovery of new gene products from *Trichoderma reesei* and *Aspergillus niger*, both of which are

widely used in industry. Again, the strategy will be to construct large insert BAC libraries, fingerprint them to assemble overlapping contigs, and follow by sequencing.

Cryptococcus neoformans causes cryptococcosis, one of the most serious fungal diseases of humans, particularly in immunocompromised populations. As HIV infection has increased, there has been a corresponding increase in cases of cryptococcosis in all areas of the world. A large-scale sequencing project is underway at The Institute for Genomic Research (TIGR) and the Stanford Genome Technology Center (SGTC), which aims to make *C. neoformans* the first basidiomycete genome to be sequenced fully.

Several other basidiomycete genome projects are described as "coming soon," including *Agaricus bisporus*, *Phanerochaete chrysosporium*, *Schizophyllum commune*, *Ustilago maydis*, *Coprinus cinereus*, *Lentinula edodes*, *Pleurotus ostreatus*, *Puccinia graminis*, and *Armillaria* species. Our best advice with regard to these projects is regularly to inspect the World Wide Web domain www.basidiomycetes.org, which is intended to provide a central location for organizing links to research and public data on basidiomycetes.

Publications and Websites Worth a Visit

Advanced Center for Genome Technology (ACGT) at the University of Oklahoma at http://www.genome.ou.edu features pages on *Cryptococcus neoformans* and *Fusarium sporotrichioides* ESTs and *Aspergillus nidulans* and *Neurospora crassa* cosmid and cDNA sequencing.

Aspergillus nidulans genetic and physical maps can be found at http://www.aspergillus-genomics.org; the site includes links to other organisms, including the *Aspergillus fumigatus* pages at the Sanger Centre.

Aspergillus website (http://www.aspergillus.man.ac.uk/index.htm) provides information on pathogenic *Aspergillus species* for clinicians and scientists (DNA sequence data, bibliography, laboratory protocols, and treatment information), and also has a separate "*Aspergillus* for patients" website (http://www.aspergillus.man.ac.uk/patients/).

Aston, C., Mishra, B. & Schwartz, D.C. (1999). Optical mapping and its potential for large-scale sequencing projects. *Trends in Biotechnology* **17**, 297–302.

Basidiomycetes website at http://www.basidiomycetes.org aims to provide a central location for organizing links to available research and public data on basidiomycetes, especially developing work on basidiomycetes genomes.

Bennett, J.W. (1997). White paper: genomics for filamentous fungi. *Fungal Genetics and Biology* **21**, 3–7.

BioKnowledge™ Library is a set of proteome databases that contain annotation and reference data for more than 50,000 proteins extracted from published research articles by expert curators. Operated by Proteome, Inc. (acquired by Incyte Genomics, Inc.) at http://www.proteome.com, the library includes the Yeast Proteome Database (YPD™) of *Saccharomyces cerevisiae*, and PombePD™ and

CalPD™, which are proteome databases for *Schizosaccharomyces pombe* and *Candida albicans*, respectively.

Birren, B., Green, E.D., Klapholz, S., Myers, R.M. & Roskams, J. (1997). *Genome Analysis: A Laboratory Manual*. Vol. 1, *Analyzing DNA*. Vol. 2, *Detecting Genes*. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

Candida albicans genome has been partially sequenced at http://sequence-www.stanford.edu/group/candida.

Cho, R.J. & Campbell, M.J. (2000). Transcription, genomes, function. *Trends in Genetics* **16**, 409–415.

Delneri, D., Brancia, F.L. & Oliver, S.G. (2001). Towards a truly integrative biology through the functional genomics of yeast. *Current Opinion in Biotechnology* **12**, 87–91.

Farman, M.L. (2001). Genome analysis in filamentous fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 91–117. Oxford University Press: Oxford, U.K.

Fungal Genomics Laboratory at North Carolina State University at http://www.cals.ncsu.edu/fungal_genomics/fromMagna/project_riceblast.html, is involved in several projects to characterize the rice blast (*Magnaporthe grisea*) genome, and has plans to study *Phytophthora* genes expressed during infection and propagation and develop new gene products from *Trichoderma reesei* and *Aspergillus niger*.

Fungal Genetics Stock Center's website is at http://www.fgsc.net. The FGSC was founded in 1960 to serve as a culture collection for *Neurospora* strains. Today, FGSC includes many thousands of *Neurospora* and *Aspergillus* cultures, and representatives of other fungi, including *Magnaporthe grisea*. In addition, the FGSC stores and supplies cloned genes, gene libraries, and EST libraries for *Neurospora*, *Aspergillus*, and *Magnaporthe*, and molecular tools for working with them. The FGSC website also provides online access to the *Fungal Genetics Newsletter*, which includes regular updates of conventional and physical genetic maps.

Genamics (http://genamics.com) is a software and web development company dedicated to improving computer tools and computer resources to provide easy access to microbial genome projects. Their GenomeSeek database contains more than 150 entries (http://genamics.com/genomes/index.htm).

Genetics, the research journal, published a special section on fungal genomics in the March 2001 issue: Vol. 157, beginning p. 933.

GenomeNet (http://www.genome.ad.jp/) is a Japanese network of database and computational services (in English) for genome research and related research areas in molecular and cellular biology. Established under the Human Genome Program of the then Ministry of Education, Science and Culture, GenomeNet is operated by the Bioinformatics Center, Institute for Chemical Research (ICR), Kyoto University. An important service is the Kyoto Encyclopedia of Genes and Genomes (KEGG) at http://www.genome.ad.jp/kegg/. KEGG is an effort to computerize current knowledge of molecular and cellular biology. It includes metabolic pathway maps and genome maps, in addition to genome sequences.

Horiike, T., Hamada, K., Kanaya, S. & Shinozawa, T. (2001). Origin of eukaryotic cell nuclei by symbiosis of Archaea in Bacteria is revealed by homology-hit analysis. *Nature Cell Biology* **3**, 210–214.

International Rice Blast Genome Consortium official website is at http://www.riceblast.org/.

Jansen, R.C. & Nap, J.-P. (2001). Genetical genomics: the added value from segregation. *Trends in Genetics* **17**, 388–391.

Kupfer, D.M., Reece, C.A., Clifton, S.W., Roe, B.A. & Prade, R.A. (1997). Multicellular ascomycetous fungal genomes contain more than 8000 genes. *Fungal Genetics and Biology* **21**, 364–372.

MapMaker and the associated manual can be obtained from http://waldo.wi.mit.edu/ftp/distribution/software/mapmaker3/. The default statistical values in this program are not suitable for use with fungal data, particularly if small progeny sample sizes (i.e., 50 or fewer progeny) are used because small sample size limits the maximum recombination value that can be reliably identified. Recommended parameter values for MapMaker when used with fungi are: LOD (log of the odds) value of 4, and θ (maximum recombination) value of 0.2.

Munich Information Center for Protein Sequences (MIPS) at http://www.mips.biochem.mpg.de/ is the bioinformatics group of the German National Research Center for Environment and Health. It is the home of the European yeast functional analysis program and the German sequencing project for *Neurospora crassa* at http://www.mips.biochem.mpg.de/proj/Neurospora/

Murray, A.W. (2000). Whither genomics? *Genome Biology* **1**, comment003.1–003.6.

The journal *Nature* operates a Genome Gateway, which is intended to be a comprehensive web resource devoted to genomics that includes original research papers, a news service, and a set of links to the most useful and informative genomics sites on the web. Access is free at http://www.nature.com/genomics/intro.html.

Nitta, N., Farman, M.L. & Leong, S.A. (1997). Genome organization of *Magnaporthe grisea*: integration of genetics maps, clustering of transposable elements and identification of genome duplications and rearrangements. *Theoretical and Applied Genetics* **5**, 20–32.

Pandolfo, M. (1999). Molecular pathogenesis of Friedreich Ataxia. *Archives of Neurology* **56**, 1201–1208.

Peterson, S.N. & Fraser, C.M. (2001). The complexity of simplicity. *Genome Biology* **2**, comment2002.1–2002.8. The electronic version of this article can be found online at http://genomebiology.com/2001/2/2/comment/2002.

Prade, R.A. (2000). The reliability of the *Aspergillus nidulans* physical map. *Fungal Genetics and Biology* **29**, 175–185.

Raamsdonk, L.M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M.C. & Berden, J.A. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology* **19**, 45–50.

Saccharomyces Genome Database (SGD™) is a database hosted at Stanford dealing with the molecular biology and genetics of *Saccharomyces cerevisiae* at http://genome-www.stanford.edu/Saccharomyces/. Along with tools to search, display, and analyze sequences, the site offers yeast chromosome maps, a guide to the literature about yeast, summaries of published information about individual genes, and tools to survey gene expression and functional analysis data, including gene disruption results, where each gene that has been deleted has its deletion phenotype displayed. The site also has a link to the section of the Virtual Library

that describes yeast model organisms (*Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans*) at http://genome-www.stanford.edu/Saccharomyces/VL-yeast.html.

Sharman, A. (2001). The many uses of a genome sequence. *Genome Biology* **2**, reports 4013.1–4013.4. The electronic version of this article can be found online at http://genomebiology.com/2001/2/6/reports/4013.

Spurr, N.K., Young, B.D. & Bryant, S.P. (1998). *ICRF Handbook of Genome Analysis* (in two volumes). Blackwell Science Ltd: Oxford, U.K.

Tait, E., Simon, M.C., King, S., Brown, A.J., Gow, N.A.R. & Shaw, D.J. (1997). A *Candida albicans* genome project: cosmid contigs, physical mapping, and gene isolation. *Fungal Genetics and Biology* **21**, 308–314.

The Sanger Centre (http://www.sanger.ac.uk/) is a genome research center founded by the Wellcome Trust and the Medical Research Council in the United Kingdom, with the purpose of furthering knowledge of genomes, particularly through large-scale sequencing and analysis. The Centre played a substantial role in the sequencing and interpretation of the human genome, but it hosts several fungal resources, including sequencing projects (see http://www.sanger.ac.uk/Projects/) for *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pneumocystis carinii*, *Candida albicans*, and *Aspergillus fumigatus*.

Thompson, A., Lucchini, S. & Hinton, J.C.D. (2001). It's easy to build your own microarrayer! *Trends in Microbiology* **9**, 154–156.

The Institute for Genomic Research (TIGR) is a not-for-profit research institute (see http://www.tigr.org) that provides databases containing DNA and protein sequence, gene expression, cellular role, protein family, and taxonomic data for microbes, plants, and humans. TIGR is involved in a large-scale sequencing project of *Cryptococcus neoformans*, and the database for that is at http://www.tigr.org/tdb/edb2/crypt/htmls/

Washburn, M.P. & Yates, J.R. III (2000). Analysis of the microbial proteome. *Current Opinion in Microbiology* **3**, 292–297.

The Whitehead Institute's Center for Genome Research is the largest public sequencing center in the world. The website at http://www.wi.mit.edu/home.html. The Center's program includes a *Neurospora crassa* sequencing project. Updated versions of this genome are being released as new data emerge and can be found at http://ww.genome.wi.mit.edu/annotation/fungi/neurospora/.

Zhu, H., Choi, S., Johnson, A.K., Wing, R.A. & Dean, R.A. (1997). A large-insert (130 kbp) bacterial artificial chromosome library of the rice blast fungus *Magnaporthe grisea*: genome analysis, contig assembly, and gene cloning. *Fungal Genetics and Biology* **21**, 337–347.

Historical Publications Worth Knowing About

Bell, G.I., DeGennaro, L.J., Gelfand, D.H., Bishop, R.J., Valenzuela, P. & Rutter, W.J. (1977). Ribosomal RNA genes of *Saccharomyces cerevisiae*. I. Physical map of the repeating unit and location of the regions coding for 5 S, 5.8 S, 18 S and 25 S ribosomal RNAs. *Journal of Biological Chemistry* **252**, 8118–8125.

Broda, P., Oliver, S.G. & Sims, P.F.G. (1993). *The Eukaryotic Genome*. Cambridge University Press: Cambridge, U.K.

Metzenberg, R.L., Stevens, J.N., Selker, E.U. & Morzycka-Wroblewska, E. (1985). Identification and chromosomal distribution of 5S rRNA genes in *Neurospora crassa*. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 2067–2071.

Schwartz, D.C., Li, X., Hernandez, L.I., Ramnarain, S.P., Huff, E.J. & Wang, Y.-K. (1993). Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science* **262**, 110–114.

CHAPTER 9

Systematics, Phylogeny, and Evolution

Revision Concepts

- *Phylogenetics* is the study of the evolutionary development of organisms in a way that allows inference of their relatedness.
- The fungal fossil record is poor and their morphological characters few, so evolutionary relationships have been derived mainly from molecular analyses.
- Macromolecules can be used to measure evolutionary change, assuming changes occur gradually and by mutation.
- Sequence comparison can reveal details that range from the level of the sequence itself to the level of population dynamics.
- Aligned sequences are processed as distance or discrete characters.
- Distance relationships are shown as branching diagrams or trees, whose branch lengths are proportional to evolutionary distances between taxa.
- Procedures for inferring trees from distance methods include UPGMA, minimum evolution, and neighbor joining.
- Discrete characters can be analyzed by the maximum likelihood method or unweighted parsimony.
- Parsimony makes few assumptions about the evolutionary process and is the best estimate of phylogeny because its only assumption is that evolutionary change is rare.
- Rooted trees identify a common ancestor, indicate the direction of evolution, and can be calibrated against geological time.
- Bootstrapping and Bremer decay indices determine how well trees are supported by data; randomization tests detect overall tree structure.
- Horizontal transfer is the exchange of genetic information within the same generation.
- Identical sequences shared by phylogenetically distant species are indicative of horizontal gene transfer.
- Population genetics is based on the measurement of allele frequencies.
- The Hardy-Weinberg model can be used to estimate allele frequencies.
- Random change in allele frequency is called *genetic drift.*

- Positive selection pressures impose a direction on the change in allele frequency, which is the basis for natural selection.
- Fitness is the capability of any particular genotype to survive and reproduce.
- The balance between sexual and asexual reproduction, and between inbreeding and outbreeding, shapes fungal populations, with the two extremes being random mating (panmictic) and strictly clonal.
- Intermediate populations arise when recombination during mating produces offspring that reproduce clonally and increase to high frequencies.
- Specific (or vertical) resistance is conferred by one or a few major genes effective against particular genetically defined races of a pathogen, whereas general (or horizontal) resistance is polygenic and effective against a wide range of pathogens.
- Host resistance genes and pathogen virulence genes interact on a gene-for-gene basis.
- An evolution study in a continuous flow fermenter of Quorn™ mycoprotein product (*Fusarium*) is the only known example designed to test selection pressures on fungi.
- Applied genetics aims to improve the production of fermented fungal products and chemicals through modification of industrial strains.
- Heterologous gene expression in fungi is important for the production of high-value therapeutics, with the main deficiency being that fungi are unable to perform posttranslational modifications such as phosphorylation, farnesylation, and glycosylation.

9.1 Phylogenetics: Inferring Evolutionary Development

Way back in Chapter 1, we made some rather amazing, yet very confident, statements such as: "... the eukaryotes emerged from their prokaryotic ancestors about 2×10^9 years ago ..."; and "... the major Kingdoms of eukaryotes have been separate from each other for the past 1×10^9 years ..."; not to mention, "... mushroom fungi first evolved about 200 million years ago." Because we have also written that the "... evolution of fungi cannot be established from a good collection of fossils ..." we will now try to explain how we can claim statements like these to be facts about the evolution of life on Earth.

The sequence of events involved in the evolutionary development of a species or taxonomic group of organisms is called the *phylogeny* of that species or taxonomic group. *Phylogenetics* is the study of phylogeny and matters relating to it in a way that allows you to infer how organisms are related. In just about any textbook of biology you will find a section describing the evolutionary march in animals from something like a sponge to a human, so it's easier for us to relate to the story of animal evolution. This

is because metazoans have left such a large and varied fossil record that the study of morphological characters in the fossils has in most cases been sufficient to provide our understanding of biological evolution. The stretch of geological time from the origin of the first metazoan to the present day, however, represents only one sixth of the total time that life has existed on Earth. Microorganisms and fungi have been evolving for a much greater amount of time, yet both their fossils and their morphological characters are fewer and more difficult to interpret than they are in animals. What we know today about the evolutionary relationships among fungi derives mainly from the application of molecular methods of analysis.

The method depends on the recognition that particular cellular macromolecules can serve as evolutionary chronometers, measuring evolutionary change. The fundamental assumption of the argument is that changes gradually occur in the sequence of the chosen molecule as a result of mutation. As a result, if two organisms are compared, the number of differences in the sequence of a molecule is proportional to the number of mutations in the DNA encoding the molecule that have been fixed in the two genotypes because the species diverged from a common ancestor. The different mutations that become fixed in different populations, of course, are the underlying cause for the biological evolution. As long as homologous molecules are studied, however, the monomer sequences of macromolecules of present-day organisms allow calculation of the evolutionary distance between the two species.

We have been careful to use the word *macromolecules* in the last few sentences because the analysis can be applied to both polypeptides and polynucleotides. It is obvious, though, that nucleic acid sequences are the most fundamental because their nucleotides are the basic units of information encoded in an organism. Other advantages are that it is becoming easier to extract and analyze nucleotide sequence information, sequence evolution is comparatively easy to model, and data sets are potentially of enormous size. Although there are many polypeptide sequences on record, as the number of laboratories collecting nucleotide sequence data steadily increases, more and more DNA sequences are added every day to several databases that are available on the Internet. GenBank [maintained by the U.S. National Center for Biotechnology Information (NCBI) at the National Library of Medicine], the EMBL data base compiled by the European Molecular Biology Laboratory, and the DNA Database of Japan (DDBJ) are the best known and most widely used. For protein sequences there is SWISS-PROT and the Protein Sequence Database of PIR-International. These databases exchange data regularly. Sequence data is evidently readily available and readily accessible.

The major applications of comparative sequence analysis are: (1) studies of gene evolution, (e.g., the origin of change in sequence, derivation of new genes and alleles, developmental studies and analysis of selection, and even

epidemiology of diseases); (2) species population studies (= intraspecific variation), including geographic variation, population genetics, ecological and behavioral analyses, gene flow, and conservation genetics; (3) interspecific studies aimed at establishing the phylogeny of species; and (4) studies in what is known as "deep time" (i.e., time scales in the hundreds of millions of years) to evaluate macroevolutionary patterns and phylogeny at taxonomic levels above the genus. To determine true relationships it is essential to choose the correct molecules for these studies. For a proper comparison, it is fairly obvious that the molecule should be present in the entire group chosen for study. In addition, the molecule must be functionally homologous, which means that it must have identical function in each organism. Molecules with different functions cannot be expected to show sequence similarities, nor can they be expected to be subject to similar selection pressures.

For effective sequence comparisons, proper alignment of the molecules being compared is crucial so that regions of sequence homology and sequence heterogeneity can be identified. Finally, and perhaps most important of all, the sequence of the molecule chosen for analysis must have a rate of change appropriate to the degree of relationship or the evolutionary distance you want to measure. Evolutionary variability of a molecule is a balance between the rate of mutational change and the constraints imposed by selection pressure on structure and function of the molecule. To measure large phylogenetic distances you must use molecules with a slow rate of sequence change; however, such a molecule would be useless to examine relationships between different geographic isolates of a particular species.

9.2 The Molecule Is the Message

Of course, a perfectly respectable reason for choosing a sequence for study is its own intrinsic interest, and/or your own interest in evolution or polymorphism of that particular gene. In this case you may need to clone and then sequence the DNA using the approaches indicated in Chapter 8. This sort of sequence comparison can provide details about the origin of alleles of enzymes where species distribution zones overlap (so-called hybrid zones), the distribution of transposable elements, differential selection, the processes of mutation and selection, and molecular, cellular, and developmental biology. Sequence variation at the species level can be used in studies of disease epidemiology, gene flow, geographic variation, hybridization, and evaluating the genetic structure of populations (called *intraspecific phylogeography*).

Sequence analysis may not be ideal for studying such intraspecific variation, however, because you can obtain detailed information about only one

or a few loci in a sequencing study. You can get information for many more loci, although it is less detailed, from studies of allozymes, DNA fingerprinting, microsatellites, RFLPs, and/or RAPDs. Interspecific studies can extend across immense spans of geological time by comparing very different present-day species. Sequences that have proved especially useful for this sort of study include the ribosomal DNA repeats and mitochondrial genomes (and chloroplast genomes in plants). Some other nuclear genes can be useful within particular groups, but you have to be careful to avoid gene duplications and pseudogenes. The comparison is only valid if the same sequence is compared, and any duplication of the sequence will make that impossible.

The nuclear ribosomal RNA genes have been particularly important for inferring phylogenies because the antiquity of the protein-synthesizing machinery means that ribosomal RNAs are large, ancient molecules; they are universally distributed, and are very constant in function and therefore quite well conserved over considerable phylogenetic distances. As a result, the degree of similarity in ribosomal RNA sequences between two organisms reliably indicates their relative evolutionary relatedness. The nuclear-encoded ribosomal DNA repeats are easily accessible, and the three ribosomal RNA sequences (see discussion of restriction mapping of the yeast rDNA repeats in Chapter 8) contain several regions of highly conserved sequences that are useful for aligning sequences obtained from different sources. Yet, they can also have sufficient sequence variability in other regions of the molecule to serve as excellent phylogenetic chronometers.

Deep time phylogenetic sequence comparisons concentrate on a comparison of the coding portions of the ribosomal genes and their RNA products. The rDNA is unusual in that between the conserved regions there are rapidly evolving regions, which are useful for examining relationships within more closely related groups; that is, relationships between closely related genera, between species, or between isolates of a single species of fungi. Eukaryotic DNA that correspond to the ribosomal RNAs consists of a cluster of 50 to several hundred copies of a repeated transcription unit, each containing 18S-5.8S-28S rRNA coding sequences arranged in that order (see Fig. 8.2). These are transcribed as a precursor molecule by RNA polymerase I. Nontranscribed spacer DNA separates the repeats, and external and internal spacer molecules are removed after transcription. The nontranscribed spacers vary in length due to variation in number of an internally repetitious sequence, so these are good candidates as molecular markers. The internally transcribed spacers (ITS) are cut to generate the mature rRNAs.

The nucleotide sequences of the ITS regions are highly variable. Although this can make them sometimes difficult to align, the polymorphism also makes them ideal candidates as molecular markers that vary rapidly enough for the study of relatively close relationships. ITS regions

are sometimes described as nonfunctional sequences; therefore, they are completely free to change by mutation without adverse selection pressure, but they do have functions during the processing of the primary rRNA transcript. ITS1 aids the processing of the 3′-end of 18S and the 5′-end of 5.8S rRNAs, and ITS2 is involved in the processing of the 3′-end of 5.8S and the 5′-end of the 28S rRNA. It is the secondary structures of the ITS regions that are important in processing, however, and this presumably accounts for their variability because maintaining a functional secondary structure is a lesser constraint on mutation in the nucleotide primary sequence than other functions of nucleotide sequences.

An important practical consideration in the exploitation of rDNA for studying relationships between fungi is the availability of universal PCR primers for the different regions of rDNA. Universal primers target conserved sequences in the rDNA and can greatly ease the task of collecting large data sets. A relative time scale for the origin and radiation of true fungi based on the 18S rDNA gene sequences was shown in Fig. 1.1, and described in words in Chapter 1. It was calibrated using the fungus fossil record to estimate the rate at which base substitutions have accumulated in fungal lineages during evolution. This rate works out to an average of 1% base substitution in any given lineage per 100 million years.

Even though there are many advantages to using rDNA sequences when attempting to infer relationships, they are not free of difficulties, nor do they solve all problems. There is a consequent trend toward inclusion of other gene loci in the data sets gathered for phylogenetic analysis. Among genes that code for proteins, the β-tubulin genes have been found to be very useful for investigating relationships between fungi at all levels, from studies of complex species groups to deep time phylogenetic investigations. There are three major tubulins. The α- and β-tubulins make up the microtubules that comprise the cytoskeleton, mitotic spindles, and flagella of eukaryotes; γ-tubulins have a role in microtubule organizing centers. Thus, like rRNA molecules, the tubulin proteins are involved in ancient eukaryotic functions and can be presumed to be ancient molecules, universally distributed, very constant in function and therefore quite well conserved over considerable phylogenetic distances. Mutations in the β-tubulin gene in fungi can produce resistance to the fungicide benomyl, which is of interest in its own right as well as a selectable marker for transformation systems. In several instances, duplicated copies of tubulin genes have been found in fungi, although duplication is more common in other organisms. There are two copies in *Colletotrichum graminicola* and *Aspergillus nidulans*, and five copies in *Epichloe* species. The β-tubulin genes belong to two lineages in ascomycetes. Their divergence appears to be very ancient and probably occurred before the basidiomycetes evolved because modern basidiomycetes have a β-tubulin sequence related to just one of the ascomycete lineages.

9.3 Inferring Relationships

We can now apply some general rules to decide which sequence or sequences we should study. To be specific, the molecules being compared must be absolutely homologous, their sequences must align with absolute reliability, and the molecule(s) must have a rate of change appropriate to the degree of relationship we want to measure. Let us assume that we've chosen a sequence to study; before we describe what to do next, maybe we should explain why we're doing it.

If you want to compare organisms, you compare phenotypic characters: that one is large, this one is small; that one is green, this one is blue; and so on. The more characters you can use, the more discriminating the comparison will be. The point about using DNA and protein sequences is that each monomer of which the polymer is composed is a character, and there are many of them. DNA sequences, in particular, provide an abundant supply of data because every nucleotide position in a sequence is a character that can exist in any one of five states. Each position could theoretically exist as any one of the four bases (i.e., adenine, thymine, guanine, or cytosine), but it might also be deleted and express a null state. In principle, therefore, using sequence data as characters for phylogenetic analysis is straightforward. In a set of sequences aligned for comparison, the characters are represented by corresponding positions in the sequences, and the states in which those characters can exist are the nucleotide or amino acid residues found in each position.

If we confine ourselves to nucleotide sequences, we can exemplify this with an example in which organism X has nucleotide T at position 45 (that is, as the forty-fifth nucleotide position in a sequence), whereas organism Y has nucleotide C at position 45 in its homologous sequence. Organisms X and Y are obviously divergent at position 45. It is unlikely (but not totally impossible) that this will be a sufficient difference to distinguish between the two organisms, so you might have to continue to examine the rest of the 30 million or so nucleotides in the genome! This simple example makes the point that in addition to positional homology, functional homology is required in the molecules used for phylogenetic analysis. In other words, nucleotides observed at a particular position in the taxa being examined should all trace their ancestry to that position in the homologous sequence of the common ancestor of those taxa.

All methods of reconstructing phylogeny start from a set of aligned sequences. The alignment is the homology that represents the shared ancestry from which the historical inferences are made. Maintaining positional homology usually requires that insertion and deletion events be postulated by inserting gaps to correspond to insertions or deletions in one or more of the sequences being compared. This is why unambiguous alignment is often difficult to achieve. Once alignment has been achieved, there are two

fundamental ways of processing the data: as distances or as discrete characters. Distance methods first convert aligned sequences into a pairwise distance matrix, then they apply a phylogenetic tree-building method to that matrix, whereas discrete methods directly consider the state of each nucleotide site in each sequence under comparison. By combining molecular biology and computer programming, this branch of genetics has inevitably developed its own vocabulary. Before we introduce you to that, we will describe the analysis of a simple example. All of these methods use computer programs, with the result that most of the processing is hidden within the algorithms from which the programs are constructed. Distance methods are easier to dissect out and explain, so we will use this approach for our example.

Distance methods usually describe the degree of relationship between two sequences, with their similarity or dissimilarity expressed as a fraction or percentage. To put this the most simply, dissimilarity between two sequences is equal to the number of aligned sequence positions containing different sequence characters (i.e., nucleotide bases or amino acids, of course) divided by the number of sequence positions included in the comparison. The numbers obtained from this simple calculation are called *distance values*; they are quantitative comparisons of two sequences, so each distance value measures a pairwise relationship. The more sequences you include in your study, the more pairwise relationships there are and the greater the length of the sequences you compare, the more opportunities there are for slight differences between them. As a result, although this may be quite routine arithmetic, its extensive and reiterative nature requires computer analyses. In this simple example we are going to compare these four short DNA sequences:

1 CGTAGACCTGAC

2 CCTAGAGCTGGC

3 CCAAGACGTGGC

4 GCTAGATGTGCC

There are just 12 characters (nucleotide bases) in these sequences, and the first difficult job of aligning the sequences has already been done, so we can proceed to the pairwise comparisons. For this we count every position in the data set in which there is a difference between the two sequences. Thus, by comparing sequences 1 and 2 we can see three differences (shown in bold and underlined: C**G**TAGA**C**CTG**A**T), and calculate a distance value of 3/12 = 0.25. Doing the same calculation for all pairwise combinations gives the distance matrix shown in Table 9.1.

The quantities shown in Table 9.1 are the raw distance values; they have to be corrected. The correction needed is a statistical correction to account for the different likelihood of the actual differences at each character site.

Table 9.1. Distance matrix of differences between four DNA sequences in pairwise comparisons.

	2	3	4
1	0.25	0.33	0.42
2	—	0.25	0.33
3	—	—	0.33

In DNA, for example, multiple changes might have occurred at any given site, perhaps two mutational events (don't forget we are looking at changes over geological time scales), with one that originally changed the sequence followed by back mutation to the original genotype, or additional forward mutations at the same site. A simple and widely used correction for DNA sequence data is based on the assumptions that substitutions follow a Poisson distribution and that all kinds of substitutions are equally likely.

Other correction routines can take into account different frequencies of transition (i.e., purine to purine, pyrimidine to pyrimidine) and transversion (i.e., purine to pyrimidine and pyrimidine to purine) mutations. If you are comparing polypeptide sequences, allowance must be made for the fact that amino acids fall into different chemical groupings (i.e., acidic, aromatic, basic, nonpolar, and polar), and some changes are usually ignored when comparing proteins. For example, aspartic acid (an acidic amino acid) in one sequence being replaced by glutamic acid (another acidic amino acid) in another sequence is judged to be a "conservative substitution" because the two amino acids are considered synonymous. Substitutions of residues in different chemical groups (e.g., leucine, which is a nonpolar amino acid, genetic code codon UUA, in one sequence, or phenylalanine, which is an aromatic amino acid, codon UUU, are considered nonsynonymous).

Phylogenetic computer programs include statistical routines that take these considerations into account and correct the raw distance values. For our example the corrected distance matrix is shown in Table 9.2; these are the numbers that describe the relationships between the sequences we are comparing. The corrected values are estimates of the total numbers of substitutions that occurred, including an allowance for all those substitutions that might have been concealed by subsequent changes at the same site. Corrected distances are usually larger than observed distances due to these multiple substitutions. The mathematics theoretically allows for downward correction, too, but this is counterintuitive as corrected distances between

Table 9.2. Corrected distance values.

	2	3	4
1	0.30	0.44	0.61
2	—	0.30	0.44
3	—	—	0.44

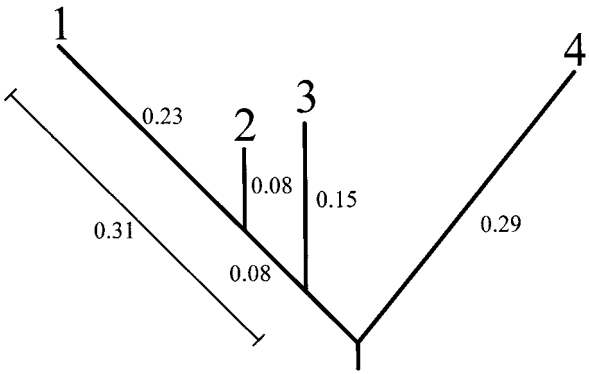


Fig. 9.1. The distance matrix phylogenetic tree for the imaginary data of Table 9.2. The tree is drawn so that branch length separating any two sequences is proportional to the evolutionary distance between them (as shown in Table 9.2). (Based on Fig. 15.11 in Madigan, Martinko & Parter (1997), *Brock's Biology of Microorganisms*, eighth edition, Prentice Hall International, Inc.)

sequences less than the observed imply that less evolutionary change took place than has been observed. This contradiction has led some workers to abandon the use of distance methods, or to use maximum likelihood methods of computing branch lengths from distances.

Phylogenetic distance relationships are typically shown as branching diagrams that are called *trees* (because that's what they look like), but we can interpret the distance matrix of Table 9.2 in words first, to explain how trees are generated from such data. Remember that the data in the table are distances, and it's quite clear that sequences 1 and 4 are the furthest apart. Sequence 4 is also far apart from sequences 2 and 3. If we convert those words into a branching diagram, it's fairly clear that a good start would be to put 1, 2, and 3 along one branch and sequence 4 on a branch by itself. Looking at 1, 2, and 3, it is also clear that sequence 1 is further from 3 than it is from 2, but then 2 and 3 are the same distance from 4. Converting that description into a branching diagram, with the distances drawn to proportional scale, gives us the tree shown in Fig. 9.1. Although simple, this is a perfectly good phylogenetic tree which shows that sequence 4 diverged first from the ancestor of (1, 2 and 3), then sequence 3 diverged from the ancestor of (1 and 2), and finally 1 and 2 diverged from each other. The total length of the branches separating any two sequences is drawn in scale to be proportional to the calculated evolutionary distance between them. This phylogenetic tree was generated so that the length of the lines in the tree would be proportional to evolutionary distances calculated from the number of nucleotide base differences between two sequences. This is the essence of the distance method.

The alternative approaches are called *discrete character methods*. The data for these approaches are the individual nucleotide sites, and the algo-

rithm compares the input data with as many trees as possible and chooses the tree that best represents the data. A maximum likelihood method uses the data to determine the probability of substitution, the relative frequencies of the four nucleotides, and the different probabilities of transitions and transversions. It then selects the tree that maximizes the probability of good fit of the data. The principle of maximum likelihood is that the explanation that makes the observed outcome the most probable occurrence must be the preferred explanation. It is important that the likelihood is not the probability that the tree is the true phylogenetic tree, but the probability that the tree has given rise to the data we analyzed. This approach requires a lot of computer power because all possible ways that the observed data could have been generated are calculated.

Parsimony is another discrete character method that creates evolutionary trees based on a systematic search among possible trees for the fewest plausible mutational steps from a common ancestor necessary to account for two diverged lineages. The minimum number of mutational changes needed to explain the original data is calculated for all possible trees, and those that require the fewest changes are said to be the most parsimonious (i.e., optimal). Parsimony optimizes the total number of evolutionary changes on the tree or tree length, where the tree length is the sum of the number of changes at each site. Unweighted parsimony treats transition and transversion mutations equally; weighted parsimony (which usually performs better) gives some changes (usually transversions) more importance in selecting the optimal tree. Phylogenetic trees based on parsimony are similar in appearance to trees based on distance, although branching order in a parsimonious tree often differs from that in a distance tree generated from the same set of sequences. Parsimony makes few assumptions about the evolutionary process, and is based on extensively studied mathematical theory accompanied by powerful computer algorithms. It is thought by many to be the best estimate of phylogeny because its only implicit assumption about evolution is that evolutionary change is rare.

In practice, the tree is itself generated by computer analysis of the data. Several different programs for sequence analysis and drawing phylogenetic trees are available. Which ever program is used, raw sequence data must be aligned first, using a sequence editor that searches the sequences, suggests alignments, and enables the user to insert gaps to take account of insertion and deletion mutations. Aligned sequences are then imported into the treeing program and the comparative analyses done. Named procedures for inferring trees by distance methods that you might encounter include unweighted pair group method with arithmetic means (UPGMA), minimum evolution, and neighbor-joining.

UPGMA is the simplest of the distance methods: it clusters data on the basis of similarity and assumes that an evolutionary clock operated, and that changes accumulated at a constant rate among all lineages (an assumption that is often not true). UPGMA is a cluster analysis method that can

be applied to all sorts of nonbiological data so it can be found as an option in most packages of statistical programs. Like UPGMA, minimum evolution and neighbor-joining infer trees from the distance matrix, but they do not assume an evolutionary clock, clustering taxa by minimizing the total distance in the tree. Minimum evolution examines every possible tree and selects the one with minimum branch lengths. This is computationally difficult because there are so many possible trees. Neighbor-joining first establishes a star tree (like Fig. 9.3) in which terminal taxa are equidistant, then temporarily takes two taxa from the star to a new node, recalculates the total distance in the new tree, returns the taxa to the star, and removes another pair to repeat the operation. This process is continued until all of the taxa are joined in a completely resolved tree with the lowest total distance. Neighbor-joining is computationally efficient and usually yields trees close to the minimum evolution tree. The neighbor-joining algorithm is included in the PHYLIP and PAUP* (PAUP-star) computer packages. Programs to perform the discrete character methods, parsimony analysis, and maximum likelihood are also included in PAUP* and PHYLIP.

9.4 Making Trees Make Sense

A variety of different names is given to the components of a phylogenetic tree, but we illustrate the more usual ones in Fig. 9.2. The terminal taxa (also called the tips, leaves, or external nodes of the tree) are the contemporary organisms (i.e., usually living species), which have been studied during the analysis. The tree is made up of branches, and the point at which a branch originates within a tree is called an *internal node* (they are also known as vertices or points). The parts of branches that connect successive pairs of nodes are internodes, links, or segments. Peripheral branches end at a tip (terminal taxon) and interior branches do not (rather, they end at the next internal node). Each node is a point of divergence between two taxa. The internodes represent lineages that evolve through time. The lengths of internodes may be made proportional to the number of sequence changes that occur in the evolving lineage (called an *additive tree*), or it may be arbitrary; it depends on the program and the preferences of the user.

When three branches are joined at an internal node (like a letter Y), then the node corresponds to a single lineage diverging into two; that is, a bifurcation or dichotomy. When there are more than three branches connected to an internal node it is called a *polytomy*. A tree that has all of its branches radiating fanlike from a single internal node is called a *star tree* (Fig. 9.3). Polytomies can represent two different situations. They usually indicate uncertainty about branching order because of ambiguous or insufficient data. The tree will become more resolved as more data are added (Fig. 9.3). On the other hand, if the data has sufficient resolution, they can also indicate that one interbreeding ancestral population gave rise to several new

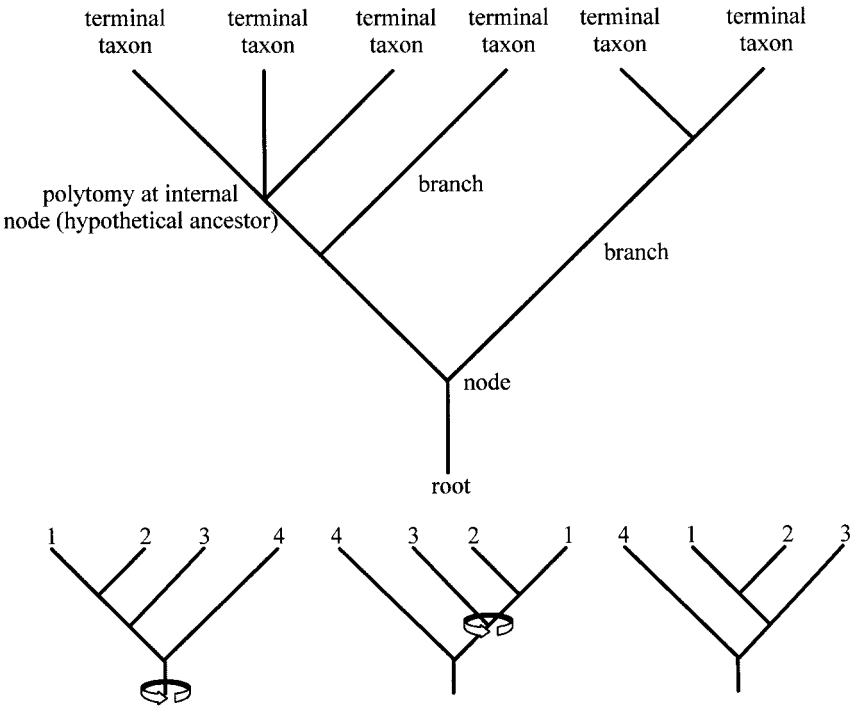


Fig. 9.2. Top panel shows the names given to the components of a phylogenetic tree. Bottom panel shows three trees that look different but are actually the same, differing only in having branches rotated around nodes.

species at the same time; that is, a true simultaneous divergence, called a *hard polytomy*. A tree that has no polytomies is described as being fully resolved, or strictly bifurcating: all of its internal nodes will be dichotomous (Fig. 9.3).

Different kinds of trees can be used to depict different aspects of evolutionary history. The most basic tree is the cladogram, which shows ancestry relationships in relative terms. Additive trees (phylograms) depict the amount of evolutionary change that has occurred along the different

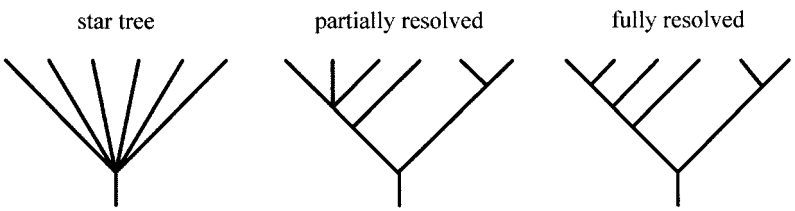


Fig. 9.3. Increasing resolution of a phylogenetic tree. (Based on illustrations in Chapter 2 of Page & Holmes (1998), *Molecular Evolution. A Phylogenetic Approach*, Blackwell Science Ltd.)

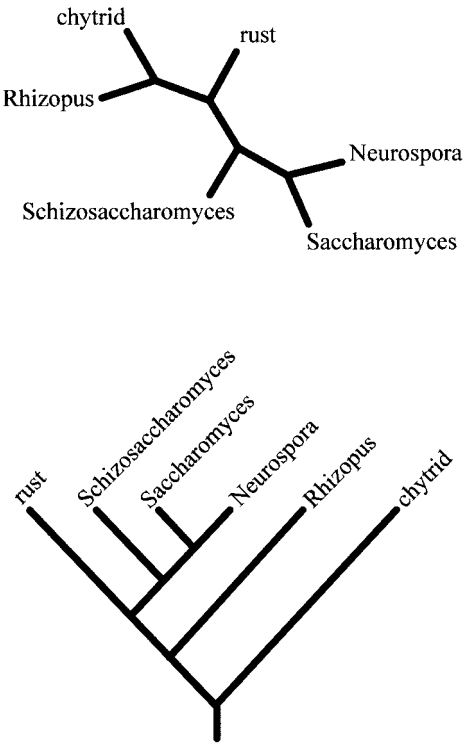


Fig. 9.4. Rooted and unrooted trees of the same data. The top, unrooted tree, shows relationships, but because it lacks a time dimension it is not possible to say what might be primitive, ancestral, or derived. For the bottom tree, the chytrids serve as the outgroup that roots the tree, and now that the direction of evolution is specified (bottom to top) it can be seen that *Saccharomyces* is more closely related to *Neurospora* than to *Schizosaccharomyces*. (Based on Fig. 2.9 in Berbee & Taylor (1999), in *Molecular Fungal Biology* (R.P. Oliver & M. Schweizer, ed.), pp. 21–77, Cambridge University Press.)

branches. The fanlike trees shown in Figs 9.2 and 9.3 are used when there are relatively few terminal taxa; however, they become difficult to interpret when large numbers of sequences are compared, in which cases dendrograms (often called *ultrametric trees*; ultrametric distances imply equal rates of evolutionary change) are used because they show the detail and depict the times of divergence more readily (like Fig. 1.1).

Cladograms and additive trees may be either rooted or unrooted (Fig. 9.4). Some computer programs used to generate phylogenetic trees produce rooted trees, whereas others produce unrooted trees. The direction of evolution (and therefore, time) is not specified in an unrooted tree. A rooted tree has a node identified as the oldest common ancestor of all the terminal taxa on the tree. A rooted tree is more informative because it indicates the direction (and, subject to calibration, the rate) of transfer of genetic information. This direction corresponds to evolutionary time and can be calibrated against geological time. In rooted trees, two branches that arise from a common node represent two descendent sister taxa, which are equal in age. The internode leading up to the bifurcation represents the evolving ancestral lineage of those two sister taxa. Unrooted trees lack this directionality, and therefore do not specify evolutionary relationships in quite the same way; however, they still show relative relationships among the taxa

analyzed as represented by the length (representing distance) of the branches joining sister taxa. Unrooted trees are mainly used to show groupings of related taxa.

Unrooted trees do not allow discrimination between ancestors and descendants: two branches that arise from a common node in an unrooted tree may be either closely related, if the node represents a recent event, or highly divergent, if the node represents an ancient event, but there is no way to discriminate between these possibilities. To find the root of the tree we must decide which two taxa are most distantly related because these must be the two groups that originated from the first divergence in the tree. We can determine the root either by using assumptions based on other information (e.g., morphology, fossils, etc.), although this can be subjective, or by using an outgroup, which is a taxon that is a near relative to, but not part of, the group of taxa in the tree. The point of divergence of the outgroup from the others defines the basal node. A third method for rooting a tree is to assume a molecular clock (i.e., to assume that a common rate of sequence change applies to all lineages in the tree). From this assumption, it follows logically that the two taxa that differ by the largest number of substitutions are the most distantly related, and that their last common ancestor is the root of the tree.

Once you have produced a tree from your phylogenetic analysis, it would be good science to ask how strongly the data support the tree. There are a number of ways of evaluating trees, but one word you will come across quite often is *bootstrapping*. You will often see bootstrap percentages on the branches of published phylogenetic trees. They show the likelihood that more data similar to the original data set would support the same branching order. In general, true resampling is impossible in most phylogenetic analyses because the terminal taxa (the present-day organisms) are individually unique (there is only one *Neurospora crassa*, for example). To work around this, bootstrapping methods create pseudoreplicate data sets. The name *bootstrapping* derives from the phrase, "pulling you up by the bootstraps," and properly implies that a difficult statistical situation is being successfully handled. The pseudoreplicate data sets are created by sampling the original character matrix at random, and with replacement, to create new data matrixes of the same size as the original. It is usual to create 500–1000 replicate data sets, each of which is used to infer trees. The frequency of appearance of a branch among all the trees generated from all the replicate data sets is the bootstrap percentage (or bootstrap proportion) of that branch. It is important to appreciate that bootstrapping is a nonstatistical measure of the support the data provides for the branches of a tree. A bootstrap percentage greater than 95% shows that the pseudoreplicate data strongly support the branch. One less than 50% indicates that more trees created from the pseudoreplicate data sets lacked the branch and consequently indicates the branch has little support.

Bremer decay indices, like bootstrap percentages, can give general support for branches in a parsimonious tree. The decay index for a branch is the number of extra character changes needed to collapse the branch into a polytomy. When a cluster is strongly supported by the data many extra changes would be required to add or remove taxa from the cluster, so it will have a high decay index. A low decay index corresponds to a branch that is less well supported by the data. Randomization tests create replicate data sets by randomizing the states for each character; this maintains character state frequencies, but removes all phylogenetic structure. The length of the most parsimonious tree from the original data set is then compared with the lengths of trees from the randomized data sets. If the most parsimonious tree inferred from the original data is not shorter than at least 95% of trees inferred from the randomized data sets, then the original data set probably lacks phylogenetic structure. Randomization tests detect overall structure; bootstrapping and decay indexes indicate the level of support for each branch.

The number and the sizes of sequences being analyzed in most molecular studies mean that the computer may be able to generate many different trees, all of which give some sort of representation of the data. It then becomes a statistical task to optimize the tree and confirm that the final tree is the best fit to the data. Because there are two ways of dealing with the data (i.e., distance and parsimony), and because there is also a need to validate the end product statistically, no single phylogenetic tree should be considered the "final judgment" on any set of relationships. Rather, each tree should be thought of as the best approximation you can reach at the moment to the true phylogeny of the sequences analyzed so far. Furthermore, gene duplications, and other events like horizontal gene transfer (see Section 9.5) and abnormal segregations, including interspecific hybridization, can confuse phylogenies based on particular genes. Thus, it is often preferable to use data representing different sequences.

Although the position is rapidly changing, most fungal classification and phylogenetic analysis still depends on morphological (i.e., nonmolecular) data. Morphological characters are unavoidably limited by the morphological simplicity of fungi, and so there are many artificial groupings in the literature that have been made on the basis of morphological similarities that are little more than coincidental. Workers in the field have debated the advantages of molecular versus morphological data in systematics, and how best to use the two types of data for phylogenetic reconstructions. Combined analyses, which use so-called total evidence, are strongly advocated by many. Others prefer only to integrate data shown to be statistically homogeneous, or to analyze different data sets separately and integrate them after morphological and molecular results have been shown to agree.

The molecular analyses that have been performed have revealed features that have an impact on the level of importance that should be given to morphological features. Indeed, the majority of characters derived from

macroscopic and microscopic analysis have evolved (or were lost and regained) more than once. For example, among mushroom fungi it seems that very similar gilled mushrooms evolved at least six times from different ancestors; therefore, possession of gills (even though this is the very the essence of being an agaric) is not a particularly helpful character for phylogenetic studies. The message that comes out of this is that for taxonomic and phylogenetic studies it is wise to use suites of characters, rather than single characters.

9.5 Horizontal Transfer of Genetic Information

Our discussion about evolutionary relationships so far assumes that DNA coding sequences change from ancestor to descendant. That is, transmission of genetic information takes place in a "vertical" direction, from one generation to the next. There is evidence, however, for transmission of genetic information from one organism to another within the same generation. This is known as the *horizontal transfer of genetic information*.

One of the first examples to be found concerned the biosynthetic pathways of the penicillins, cephalosporins, and cephamycins, which are β-lactam compounds synthesized by a wide range of microorganisms, including fungi, actinomycetes, and Gram-negative bacteria. Molecular comparisons of gene sequences and gene organization in these different microbes revealed them to be so similar that the idea that they evolved independently seemed less likely than the proposition that in some way the genes encoding the biosynthetic enzymes might have been transferred from bacterial β-lactam producers to filamentous fungi about 850 to 950 million years ago. There are several other examples that arise from DNA sequence comparisons: evidence that an intron in an angiosperm mitochondrial gene arose recently (in geological terms) by horizontal transfer from a fungal donor; horizontal transfer of a mitochondrial plasmid from the discomycete *Ascobolus immersus* to the pyrenomycete *Podospora anserina*; and the distribution of the transposable element *mariner*, which has been found in many species of dipteran flies, several other groups of arthropods, in platyhelminth worms, and a phytopathogenic fungus. At the level of the metabolite product, the hallucinogen compound bufotenin is produced by the toad *Bufo* and by a species of the mushroom *Amanita*.

These examples of similar or identical sequences being shared by phylogenetically distant species are unlikely to be relics of some ancient shared ancestor or instances of parallel evolution. Instead, it seems that the powerful mechanism of horizontal transfer is the cause. By horizontal gene transfer we mean a process by which genetic information of one organism is incorporated into the genome of another organism that might be the same or a different species. Horizontal transfer is acknowledged to be of major importance in the evolution of prokaryotes. Horizontal transfer of fungal

mitochondrial genes has been demonstrated, and although there is very little direct experimental evidence for horizontal transfer of fungal nuclear genes, there is plenty of circumstantial evidence.

A remarkable property of filamentous fungi is that genes involved in secondary metabolism tend to be clustered, usually being located less than about 2 kb from each other. Examples of secondary metabolite clusters in fungi include genes involved in toxin biosynthesis, ergot alkaloids, gibberellins, lovastatin, penicillin, sterigmatocystin, aflatoxins, and trichothecenes. Most fungal genes for other biosynthetic pathways conform to the usual eukaryotic model of being dispersed throughout the genome. The secondary metabolite clusters include regulatory genes and, where appropriate, genes conferring autoresistance, so they appear to be self-contained. Aside from the selective advantage that the metabolite pathway may confer on the organism, it has been suggested that clustering confers selective advantage to the cluster itself because the genes depend on horizontal gene transfer for their dispersal and survival. In other words, it is a selfish cluster made up of selfish genes in the sense developed by Richard Dawkins, who views living organisms as survival machines designed to preserve the selfish molecules we know as genes.

The way in which horizontal gene transfer is accomplished is unknown. There does seem to be a correlation with organism lifestyle. For example, parasitic bacteria have acquired more eukaryotic genes than free-living bacteria, and much more gene exchange has occurred between organisms that occupy the same ecological niche (like thermophiles) than between organisms with different ecological preferences. The occurrence of hyphal anastomoses may provide a direct route in fungi because even an incompatible reaction may enable gene exchange to occur occasionally.

Some genome comparisons, however, have revealed major gene transfers between bacteria, archaea, and eukaryotes, so it is clearly not dependent on hyphal behavior. In one study, comparison of complete genomes of 15 bacteria, four archaea, and one eukaryote found 37 cases of interkingdom gene fusions (IKFs), which are genes coding for proteins that consist of domains originating from different primary kingdoms. These, of course, are rather different from the selfish clusters we have just been describing, but at the time of writing we do not have a sufficient number of fungal genomes sequenced to make a search for secondary metabolite clusters. IKFs are thought to have evolved via horizontal transfer of a gene coding for the alien domain (or for a larger protein that contains that domain) followed by recombination with a native gene.

9.6 Genes in Populations

Phylogenetics deals with species or taxonomic groups of organisms or populations, and because all evolutionary changes start with changes in populations, this is about the right time to look at the population genetics

of fungi. We will start, though, by establishing the background to population genetics because it has been established for other, and importantly, *diploid*, eukaryotes. The foundation of population genetics is the frequency of alleles, the alternative forms of a gene at a given locus, and the changes in allele frequency over time. Whereas macroevolution deals with the origin of new species and the fate of species through geological time (i.e., pretty much what we have been discussing so far), microevolution covers change in allele frequencies over recordable time within a species.

The coexistence of two or more alleles at a locus is a polymorphism, and the allele frequency is the number of copies of one of the different alleles at the locus expressed as a fraction of all of the alleles at that locus. If there are two alleles, p and q, then the total number of alleles (the gene pool) = $p + q$ and the frequency of p is given by the equation $p/(p + q)$, whereas the frequency of $q = q/(p + q)$. Allele frequencies can be measured directly in *haploid* organisms because there are no complications with dominance and the allele phenotypes are directly expressed. In populations of diploid organisms, however, for a single locus with two alleles under random mating, the genotype frequencies fall into the Hardy-Weinberg equilibrium of $p^2 + 2pq + q^2$. This equilibrium has been the foundation for most theoretical studies in the established field of population genetics and must be the yardstick against which we compare fungal population genetics.

The assumptions that underlie the Hardy-Weinberg equilibrium are:

1. it deals with a single, autosomal locus with two alleles and normal (Mendelian) meiotic segregation
2. randomly mating (or panmictic) population
3. no mutation
4. infinite population size
5. no inward or outward migration
6. the age structure of the population must be in discrete generations
7. all phenotypes are equal and contribute equally to the gene pool (i.e., no selection permitted)

In addition, only one generation of random mating is necessary to achieve Hardy-Weinberg equilibrium, regardless of the starting genotype frequencies. Remember, also, that this is an *equilibrium*; there is no change in allele frequencies over time (i.e., no evolution) at the Hardy-Weinberg equilibrium (because of the "no mutation" assumption). The Hardy-Weinberg model has a number of important applications. For example, the equilibrium equation can be used to estimate the frequency of individuals in a population that are heterozygous for a deleterious recessive allele. Knowing this can be important in medical genetics because it represents the number of "carriers" in the population and can aid genetic counseling. Consider the gene that causes phenylketonuria, which is a human metabolic disorder that results in severe mental retardation if not properly treated. The recessive allele causes the disorder and is symbolized a; the dominant allele, with

normal metabolism, is symbolized *A*. Suppose that in a particular human population the frequency of individuals suffering phenylketonuria is 1 in 10,000. These are the homozygous recessives, but the homozygous dominant, *AA*, and heterozygous, *Aa*, individuals are both clinically normal and cannot be distinguished directly.

We can use the Hardy-Weinberg equilibrium equation ($p^2 + 2pq + q^2$), however, to estimate the allele frequencies, and from them the proportion of heterozygotes (carriers) in the population. We know the frequency of *aa* is 1/10,000 = 0.0001 = q^2, and it follows that the frequency of allele *a* = $\sqrt{0.0001} = 0.01$. Remembering that [frequency of *a*] + [frequency of *A*] = 1; then the frequency of *A* must = 1 − 0.01 = 0.99. Now we can use these two allele frequencies to estimate the frequency of *Aa* heterozygotes in the population. In the Hardy-Weinberg equilibrium equation the heterozygotes = $2pq = 2(0.01 \times 0.99) = 0.02$. Approximately 2% of the population, therefore, is a carrier of the phenylketonuria allele. This example uses the frequency of one genotype to calculate allele frequencies on the assumption that the population complies with the criteria of the Hardy-Weinberg equilibrium (i.e., random mating, no mutation, no inward or outward migration, etc.). If you have a means of measuring the frequencies of the alleles (or at least the genotypes) in the population directly, then you can test whether or not those frequencies comply with the Hardy-Weinberg equilibrium equation. If they do not, then one or more of the criteria underlying the Hardy-Weinberg equilibrium must have been violated.

It is violations of the assumptions of the Hardy-Weinberg model that change allele frequencies and, consequently, drive evolution. Those violations include mutation (generally an infrequent occurrence) and recombination. Evolutionary advantages of recombination include production of novel genotypes that could allow organisms to adapt quickly to changing environments. In addition, advantageous gene combinations arise more rapidly by recombination than in populations that reproduce only by asexual means because these can only produce novel genotypes by successive mutations. Finally, recombination enables the purging of deleterious mutations. Recombination does not directly change allele frequencies; rather, it greatly enhances the genetic variation produced by mutation by creating the gamete genotypes that expose recessive alleles to selection. Because of these features, sexual reproduction allows faster response to changing environments and easier removal of any deleterious mutation.

Linkage of genes can cause a violation of the Hardy-Weinberg model, called *linkage* (or gametic) *disequilibrium*. Assume we have two loci each with a pair of alleles (Aa and Bb). There are four different combinations for these alleles: AB, Ab, aB, and ab. If recombination occurs freely, then, on average, there will be an equal proportion of each combination in the population: this case is called linkage equilibrium. On the other hand, if the genes are reasonably close together, their alleles will not be randomly assorted in any one mating. Rather, particular alleles (the parental combi-

nations) will be inherited together more frequently. This is linkage disequilibrium. With random mating in a large population, mating will be taking place between individuals with all combinations of alleles of these two genes. Thus, linkage equilibrium between genes is eventually attained.

Mutation can cause disequilibrium, and it can be maintained by selection because selection can increase the frequency of an advantageous allele and other nonselected components of the genotype with which it shows disequilibrium. This process is known as *hitchhiking*. Selection for phenotypes controlled by the interaction of genes at more than one locus, also called *epistatic selection*, can also maintain disequilibrium. An extreme example is provided by the gene-for-gene relationships between fungal pathogens and their hosts (see Section 9.8). In these systems, host cultivars with multiple resistance genes select for fungal pathotypes with multiple compatible avirulence genes (and vice versa). As a result, both the resistance genes in the plants and the avirulence genes in the fungi show linkage disequilibrium. In the absence of a selection process to maintain linkage disequilibrium, recombination in a sexually reproducing organism will determine the rate of dissipation of the disequilibrium caused by physical linkage. High rates of recombination mean the population can reach equilibrium in a few generations; with low rates of recombination, the process may take millions of years.

An important, and inevitable, violation of the Hardy-Weinberg assumptions is that there is no such thing as infinite population size. Finite population size alone ensures that evolution will occur through sampling error, and the change in allele frequency that results from a restricted sample of the population being involved in reproduction. This gives rise to random changes in allele frequency, which is known as *genetic drift*. Pure genetic drift is truly random and has no direction; in particular, there is no tendency to return to ancestral allele frequencies. Changes caused by genetic drift accumulate with time, and drift can cause both loss and increase of genetic variability within a population. Genetic drift is an evolutionary force that can alter populations through time, and emphasizes that the Hardy-Weinberg equilibrium does not hold exactly for any finite population. The amount of evolutionary change associated with this sampling error is inversely related to population size; the larger the population, the less the allele frequency will change. Hence, genetic drift is most effective as an evolutionary force when the population size is small.

Genetic drift is also most effective with neutral mutations because an allele subjected to positive selection pressure imposes a direction on the change of allele frequency. This, of course, is the basis of natural selection, and it relates to the fitness of the gene—the capability that any particular genotype has to survive and reproduce. In molecular evolution, fitness is usually expressed in terms of a selection coefficient, s, which is a measure of the reduction in fitness relative to the best genotype in the population. New mutations are usually compared with the currently generally existing

genotype (the "wild type"). In most cases mutations that occur now are deleterious (i.e., they have lower fitness). Such mutations will be removed from the population (i.e., the frequency of the mutant allele will be reduced) by adverse or negative selection. If a new mutation does prove to have a higher fitness than other variants in the population, then it may increase in frequency due to positive selection.

Mutation, selection, population size, and migration are all likely to have similar influences in diploids and haploids alike (aside from any differences in mathematics between the two), so we might expect that many of the features of the Hardy-Weinberg model *could* be applied to a population of haploid fungi. For a haploid organism, though, the equilibrium genotype frequencies of the gene pool = $p + q$, and allele frequencies can be measured directly from the vegetative phase. In most diploid or dikaryotic fungi, allele frequencies can be estimated using the Hardy-Weinberg equilibrium. Codominant markers are best for such analysis so that heterozygotes can be identified. Isozymes and RFLPs are ideal, but RAPDs are less useful because they are usually dominant.

Most fungi have enormous potential for asexual reproduction, and this is likely to cause deviation from the Hardy-Weinberg model. Sexually reproducing populations are more genetically diverse than asexual populations of the same species. For example, populations of the oomycete *Phytophthora infestans* deviate greatly from Hardy-Weinberg expectations in the strictly asexual population in the United States, but in central Mexico, where the sexual stage is common, genotype frequencies are in Hardy-Weinberg equilibrium. In *Erysiphe graminis* f. sp. *hordei*, the barley powdery mildew fungus, numerous asexual cycles similarly enhance and maintain linkage disequilibrium in most populations, and, in addition, epistatic selection by host plant resistance genes also promote linkage disequilibrium.

9.7 Genes in Fungal Populations

Consider a representative fungal mycelium. Hyphal growth is indeterminate, so it can continue to grow within its substrate and colonize new substrates for an indefinite length of time. At any one point in time it is likely to be engaged in various methods of reproduction involving both meiotic and mitotic nuclear divisions producing many millions of sexual and asexual spores. It may also be producing mycelial propagules like stromata and sclerotia. Any of these spores and similar propagules might be immediately dispersed to engage in immediate competition for colonization of new substrates in some distant place, or may lie dormant for long periods to reemerge and engage in competition at some distant time, or both. Our imaginary representative fungal mycelium is clearly competing with other members of the population by dispersing its own genotype and challenging other genotypes to tests of fitness. At the same time, it is sheltering its geno-

type in dormant propagules so that the challenges can be renewed against some future population.

Ask yourself, where is the individual in this? The whole mycelium, or one unicellular and uninucleate spore? Which should you count for a population census? We have described the self–nonself recognition involved in the vegetative compatibility response (Chapter 2) that determines mycelial individuality and thereby restricts heterokaryosis. Fungi evidently have made a considerable investment in a genetic mechanism able to determine the specificity of somatic recognition. Perhaps we should follow that lead and identify genetic clones as the individual units of fungal populations. In natural populations clones can be identified as multiple-gene genotypes, which are encountered too frequently to have arisen by chance through sexual reproduction. This approach has shown that the clonal structure of fungal populations is as varied as might be expected in a large and successful kingdom of organisms.

The distribution of clones has been studied particularly in plant pathogenic fungi, and there are examples of very local dispersal (e.g., *Septoria tritici*, which causes wheat leaf spot), clonal dispersal over a continental land mass (e.g., *Sclerotinia* clones dispersed across North America), and even global dispersal (e.g., clonal lineages of *Phytophthora infestans*, (which is one of the Oomycota with a life style like a mycelial fungus, having a worldwide distribution). Other fungi are highly territorial, but this can be expressed on very different spatial scales. We have already described (Section 2.1) how some clones of *Armillaria* species can extend over many hectares, but this contrasts with other fungi, like *Lentinula* and *Heterobasidion*, which have territories only a few centimeters in extent. Many competing individuals may colonize the cut surfaces of individual tree stumps, arriving as airborne basidiospores with a strong somatic incompatibility system. In the case of *Heterobasidion*, which causes a root rot disease of trees, a single individual predominates in the lower portions of the root system of the stump. This individual might represent a resident clone that caused the initial disease, and which is being faced with competition from adventurous immigrant clones at the surface of the cut stump. Understanding the size of fungal territories is important. There are many circumstances in which some attempt must be made to manage populations (e.g., the control of pests or pathogens, the production of crops, or the conservation of biodiversity), and we must be aware of the appropriate scale at which to operate.

Taking into account the dispersal patterns of fungal populations can make an important point. At small scales, biological factors may be most strongly correlated with genetic variability. These factors include inter- and intraspecific competition, age and species composition of other, interacting, organisms, and type and severity of habitat disturbance. At larger scales the major influences on genetic variability may be physical factors such as climate, rainfall patterns, exposure, and elevation. Understanding evolu-

tionary processes critical to the maintenance of genetic diversity in fungal populations requires appreciation of how the fungal lifestyle is modified to different spatial scales by different life history strategies. This is because patterns of genetic isolation and estimates of gene flow between populations are calculated by using estimates of variance in allele frequencies within and between samples of the population(s). Those samples must cover different random breeding units (also known as genetic neighborhoods) if the estimates are to be accurate. As a result, knowing the size of the areas in which random mating occurs is essential to deciding what sampling strategy should be used to get more information about the population structure.

Fungal population structure cannot be predicted from the life history of the fungus because sexual and asexual reproduction both contribute to the structure of the population. It is the balance between the two, combined with the balance between inbreeding and outbreeding, that actually shapes the population. An example is *Sclerotinia sclerotiorum*, which is a pathogen of several plants. This fungus goes through a sexual cycle every season, but it nevertheless has a highly clonal population structure. The reason is quite simply that the fungus self-fertilizes, with the result that progeny of the meioses are genetically identical. The majority of plants and animals reproduce primarily through the sexual cycle, but asexual reproduction is not at all unusual among microorganisms. Because offspring are genetically identical to each other and their parent (except for new mutations), clonal populations are distinguished by the widespread occurrence of identical genotypes and absence of recombinant genotypes.

In contrast to this circumstance is a population in which sexual reproduction predominates. The independent assortment of chromosomes and recombination between homologous chromosomes during meiosis in this case produces a population structure featuring relatively high levels of genotypic diversity and (once any linkage disequilibrium has been dissipated) random associations between alleles at different loci. In such a population, genotype frequencies can be predicted from the allele frequencies at each locus. In terms of discussion based on a Hardy-Weinberg model, a clonal population exhibits linkage disequilibrium on a grand scale. Complete genomes, made up of independent sets of genes, are distributed through the population as an undivided package. Of course, a clonal population cannot approach Hardy-Weinberg equilibrium because it is not undergoing mating, let alone random mating.

Thus, it is relevant to ask whether we can examine populations found in nature to establish if their structure is consistent with the occurrence of random mating. The two extremes of the spectrum are random mating (panmictic) versus strictly clonal population structures. Intermediate population structures can arise if recombination during mating produces one or several successful offspring that reproduce clonally and increase to high frequencies. This intermediate model is applicable to many fungal plant pathogens.

A large number of these have annual sexual cycles and asexual epidemic phases, whereas others (e.g., the chestnut blight fungus, *Cryphonectria parasitica*) have selfing and outcrossing in the same population, which is something typical of many plant pathogens. Because selfing in haploids is equivalent to clonal reproduction, the population structure consequently deviates from that expected with random mating. Another way in which the population may be structured so as to prevent full random mating is when reproductive isolation occurs between subpopulations, although recombination is frequent within each subpopulation. Such a situation arises in plant pathogens when particular races of the pathogen are restricted to particular host cultivars: it has been observed in the maize pathogen *Cochliobolus carbonum* and *Cryphonectria parasitica*. Other populations of *C. parasitica* are clonal; clearly, different population structures can be found within the same species.

Direct assessment of the potential for sexual reproduction in natural population requires both laboratory and field studies to search for sexual structures and to assess the range of genotypes and frequencies of alleles. Presence of sexual structures in nature is an indication of the potential to reproduce sexually, but no proof of random mating. Sexual fruiting bodies may be common in a population in which significant disequilibrium exists because of selfing and/or frequent rounds of asexual reproduction. In some species, the simplest way to detect random mating is to survey the mating-type alleles. If populations are randomly mating, the mating-type alleles should have equal frequency. Sexual populations have a greater diversity of genotypes, and there are several ways to test the random mating hypothesis statistically.

Phylogenetic analysis of molecular markers can also be used to test for random mating. This approach has been used with the human pathogenic fungus that causes valley fever, *Coccidioides immitis*. The expectation is that a reliable phylogenetic tree will be obtained if the population is clonal, but not for recombined genotypes because the randomizing effect of recombination removes the phylogenetic consistency. Data for *C. immitis* lack phylogenetic consistency, implying that the population structure is driven by recombination despite the fact that there is no known sexual stage in *C. immitis*. The occurrence of recombinant genotypes is another simple way of detecting recombination in a population. New DNA fingerprint and allozyme genotypes that have arisen in *Phytophthora infestans* and in *C. immitis*, as well as the nuclear fingerprint markers in *Sclerotinia sclerotiorum*, provide examples of recombination in the most common clones of these fungi.

Study of population structure is a necessary prerequisite to understanding how organisms interact in populations and with their environments and other organisms. The emphasis so far has been on plant pathogens and their evolutionary interactions with their host.

9.8 Genetic Variation in Hosts and Pathogens

Vertical or horizontal resistance categorizes the responses of host plants to fungal diseases. Vertical resistance is generally observed in annual crops that are not vegetatively propagated, and it usually depends on one or a very few pathogen-specific genes in the plant; if the resistance is overcome the crop fails. Horizontal resistance (probably the most common in nature) is seen in perennial as well as vegetatively propagated annual crops; it is polygenic and tends to be broadly based and nonspecific. Tolerance of a disease is defined as the ability to produce a crop in spite of infection. Defense of a plant against pathogens may therefore involve specific resistance conferred by one or a few major genes, which are effective against particular, genetically defined races of a pathogen, or may involve a general polygenic resistance, which is effective against a wide range of pathogens. These are the defense strategies that the fungal pathogen must overcome to attack the host.

In the mid-twentieth century, combined studies on the inheritance of specific resistance in flax and the virulence factors of its fungal pathogen, flax rust, introduced the concept that host and parasite genes both played a role in the determination of whether or not a resistance reaction would be observed. The concept is that the expression of resistance by the host is dominant, whereas the expression of nonvirulence by the parasite is dominant (i.e., virulence is a recessive character in the fungus) and, specifically, that each individual resistance gene in the host interacted with a corresponding single gene in the pathogen. This is the gene-for-gene concept. In the simplest gene-for-gene interaction the reaction between each pair of resistance and susceptibility alleles in the host with its matching pair of virulence and avirulence alleles in the pathogen is fairly simple, but with several such interactions taking place, the overall relationship between a host and its pathogen is complex. Twenty-nine separate host resistance factors have been identified in flax, with each having a complementary virulence factor in flax rusts. Similar complementary major gene interactions between hosts and pathogens are known or suspected for more than 25 different host–pathogen combinations.

Table 9.3 illustrates the interaction between one gene locus in the host plant and one locus in the pathogen. The host alleles are symbolized *R* for resistance and *r* for susceptibility, whereas the corresponding pathogen alleles are symbolized *V* for avirulence and *v* for virulence. A two-factor interaction is shown in Table 9.4, with the array simplified to distinct phenotypes using the dominance relationships. In such a multifactor interaction, the host is resistant to the pathogen if a matching pair of (host) resistance and (pathogen) virulence alleles occurs at any one of the gene-for-gene loci. A model of the molecular mechanisms underlying these interactions has been deduced, although there is not much supporting data available yet. Gene-for-gene resistance often causes a hypersensitive

Table 9.3. Single factor gene-for-gene interaction.

Fungus pathogen genotype*	Host plant genotype*		
	RR	*Rr*	*rr*
VV	Resistant	Resistant	Disease
Vv	Resistant	Resistant	Disease
vv	Disease	Disease	Disease

* The host alleles are symbolized *R* for resistance and *r* for susceptibility, whereas the corresponding pathogen alleles are symbolized *V* for avirulence and *v* for virulence.

response, which is due to the accumulation of defensive chemicals in the plant tissue surrounding the point of infection. This serves to confine the pathogen and limit its proliferation.

Hypersensitivity is allele-specific and suggests that the (dominant) resistance phenotype corresponds to production of a particular molecular product that attacks the invading pathogen. Because pathogen avirulence is also the dominant allele in the fungus it is likely that this also corresponds to production of a particular molecule, which is not produced by the corresponding virulence allele. As a result, the panels of gene-for-gene interactions are interpreted as recognition reactions between the host and the pathogen. The avirulence alleles of the pathogen somehow label it "pathogen"; the resistance alleles of the host give the plant the ability to recognize and constrain the fungus, but it only needs one of the several potential recognition events to be successful to achieve that resistance. For the fungus, future success as a pathogen depends on avoiding recognition (i.e., on removing all of its labels so it can blend into the background, like a thief in the night). For the plant, future success depends on more and more sensitive surveillance, so that however dark the night, the thief will be seen. Gene-for-gene interactions lock the host and pathogen into a coevolutionary conflict between balanced polymorphisms.

Gene-for-gene interactions are also a target for applied genetics manipulation. Extensive use is made of major genes for resistance in disease

Table 9.4. Two-factor gene-for-gene interaction.

Fungus pathogen genotype*	Host plant genotype*			
	*R*1-, *R*2-	*R*1-, *r*2*r*2	*r*1*r*1, *R*2-	*r*1*r*1, *r*2*r*2
*V*1-, *V*2-	Resistant	Resistant	Resistant	Disease
*V*1-, *v*2*v*2	Resistant	Resistant	Disease	Disease
*v*1*v*1, *V*2-	Resistant	Disease	Resistant	Disease
*v*1*v*1, *v*2*v*2	Disease	Disease	Disease	Disease

* The host alleles of factors 1 and 2 are symbolized *R* for resistance and *r* for susceptibility, whereas the corresponding pathogen alleles are symbolized *V* for avirulence and *v* for virulence. To reduce the size of the table, homozygous and heterozygous dominants (e.g., *R*1*R*1 and *R*1*r*1) are combined (e.g., shown as *R*1-).

control strategies for agricultural crops. The applied approach depends on the reproductive strategies of the pathogen. As discussed in the previous section, pathogens with an exclusively asexual mode of reproduction differ markedly from those with periodic cycles of sexual reproduction. In clonal populations the variety of different pathogen genotypes in the population will be restricted. In contrast, sexual reproduction produces a much wider diversity of pathogen genotypes, including new virulence combinations as a result of recombination. Even in these populations, though, linkage disequilibrium will maintain the frequency of virulence alleles that are not subject to direct selection.

Hitchhiking selection may develop rapidly during epidemics as the number of asexual generations increases after the round of mating. Disease resistance genes used in breeding agricultural crops originally arose following coevolution of the pathogen (which is now the crop disease) with the ancestors of what is now the crop plant. The gene-for-gene relationship, particularly in rusts and mildews, is also found in diseases of wild plants in nature. Indeed, wild species related to crops are the commonest sources of the genes used by plant breeders to develop race-specific resistant cultivars of crops. Well-studied natural host pathogen systems include crown rust of wild oats (*Avena* spp.) and powdery mildew of groundsel (*Senecio vulgaris*).

In medical mycology, the gene-for-gene model established for plant–fungus pathogen interactions has not been observed, but the feeling is that the correct way of expressing the situation is that it has not *yet* been observed. The crucial feature in plant pathology is that mutation of avirulence genes in the fungus leads to escape from pathogen-recognition mechanisms and hence the onset of disease. Mutants in which virulence is increased have not been reported in fungi pathogenic to humans, but hypervirulent mutations are not impossible. Medically important fungal pathogens fall into two categories. Those that cause invasive disease like *Coccidioides immitis*, *Histoplasma capsulatum*, *Blastomyces dermatitidis*, *Paracoccidioides brasiliensis*, and *Cryptococcus neoformans*, together with some fungi that cause skin diseases, are all fungi that are able to infect healthy, immunologically intact people (although most instances of cryptococcosis caused by *C. neoformans* arise in immunologically compromised patients). These fungi contrast with species that are normally unable to infect healthy individuals, although they can cause infections (often fatal) in patients with suppressed immune function; this group includes *Candida albicans*, *Aspergillus fumigatus*, *Rhizopus* species, and *Fusarium* species. Molecular studies of virulence in these pathogenic fungi are revealing complex interactions between each fungus species and the human host.

Studies of the molecular aspects of virulence have revealed many features that contribute to pathogenicity. The only common theme is that no fungus depends on a single molecule for virulence; rather, virulence requires expression of several genes. Current studies aim at distinguishing between gene products that enable infectivity because they serve a house-

keeping function that is equally essential to the fungus outside the host, and those that are expressed uniquely during infection.

9.9 Evolution in Captivity: Natural and Artificial Selection

In the first half of the twentieth century, the fruit fly *Drosophila* was used in critical experimental studies of evolution. Populations of the fly were isolated in "evolution cages" and subjected to various selection pressures to test (and, incidentally, confirm) the basic tenets of Darwinian "theory." These studies showed that evolution doesn't need millions of years to happen, but is open to experiment and is relevant to the applied genetics of crop species. Very little of this lead unfortunately, has been followed by mycologists. A considerable amount has admittedly been learned about the population genetics of fungi, including natural populations, from the efforts of plant pathologists, as summarized earlier.

We unfortunately, and remarkably, know of only one sequence of studies that takes up the challenge of the fungal equivalent of the *Drosophila* evolution cage to study evolution in the real world: the continuous flow fermenter. Such a fermenter is used to grow the only successful fermenter-grown fungal food on the market, the mycoprotein Quorn™, which is the mycelium of a species of *Fusarium*. The marketability of the product centers on the virtues of its filamentous structure, which enable it to be processed to simulate the fibrous texture of meat. Coupled with the inherent nutritional value of fungal biomass, this permits the product to be sold as a low-fat, low-calorie, cholesterol-free health food and a healthy alternative to meat. The fungus is grown in a 45 m tall airlift fermenter that is used in continuous culture mode; medium is added continuously, and mycelium + spent medium is harvested constantly at a rate equal to the production of new hyphae. Given that the sparsely branched filamentous nature of *Fusarium* mycelium is crucial to the creation of the meatlike texture of Quorn™, it is essential that the filamentous character of the fungus be maintained. A variety of morphological mutants arise, however, during long-term (800–1000+ hours) cultivation in the continuous flow fermenter. In Petri dish cultures the morphological mutants (known as "colonial mutants") form highly branched hyphae in colonies that have slower radial growth rates than normal. Although colonial mutants have a reduced radial growth rate, their increased branching results in an increased rate of biomass production, or specific growth rate. In the fermenter, though, the colonial mutants have a positive selection advantage and come to dominate the culture. The mycelium of colonial mutants unfortunately does not have the right texture to make Quorn™, with the result that when they appear the fermenter must be closed down, cleared out, cleaned, sterilized, and recharged. This is an extremely expensive process with a fermenter of 155,000 l capacity. Why, then, is it necessary?

When the fermenter is first charged with a culture, it is, of course, a pure culture of uniform genotype; however, when it has grown to full capacity it will be a *very* large population of fungal nuclei (155 m^3 of mycelial suspension!), and a great many spontaneous mutations must occur among so many nuclei. During Quorn™ mycoprotein production there is no way that the mutations can accumulate unless they have some selective advantage because the culture is being constantly removed in the continuous flow of medium through the fermenter, which is operated as a glucose-stat (i.e., a chemostat in which glucose is the limiting nutrient). All the morphological mutants that become predominant in the fermenter form mycelia that are more highly branched than the wild type. They are genetically recessive to the wild type and belong to any one of three complementation groups (presumably, three functional genes).

Colonial mutants replace the parental strain (that is, they have a positive selection coefficient) in continuous flow cultures at both high and low dilution rates using nutrient limitation for glucose, magnesium, ammonium, and sulfate, suggesting that none of these aspects of nutrition are involved in the selection process. It is interesting that the parental strain replaces colonial mutants in batch cultures (that is, colonial mutants have a negative selection coefficient in batch culture). Testing the culture for inhibitor-resistant conidia reveals the periodic appearance of new mutant populations, each being indicated by the appearance of peaks in the concentration of unselected mutations, at intervals of about 124 hours (about 30 generations). These new mutant populations appear before the colonial mutants become evident in population samples. This shows that populations of filamentous fungi undergo regular bursts of mutation and, in the case of the Quorn™ production procedure, become heterogeneous before the onset of the continuous flow mode in the culture.

A culture that was assayed in this way for about three months, a time span that is equivalent to 530 generations, experienced 15 adaptive peaks at intervals averaging 135 ± 10 hours. A colonial mutant appeared after 648 hours of cultivation and increased in concentration to form eventually 92% of the population. The appearance of colonial mutants was delayed by switching the pH of the culture between 4.8 and 6.6 at intervals corresponding to the adaptive peaks, indicating that the population evolved more slowly under oscillating pH than under constant pH. On the other hand, highly branched mutants were detected much earlier in cultures supplemented with mycological peptone (peptone is partially hydrolyzed protein), although they failed to displace the sparsely branched parental population. The nutritional conditions clearly influence the evolution and selection of colonial mutants but not in a very specific way. The best that can be said at the moment is that growth conditions in the Quorn™ production plant favor mutations that confer general growth rate advantages.

It is not yet known, however, why *colonial* mutants in particular have such a selective advantage that they can come to be the predominant mor-

phology in such large cultures. The selection pressure, which is exerted in favor of colonials in continuous flow fermenters, is most likely to be biophysical rather than biochemical. The highly branched growth form will have a larger surface area to mass ratio, which could affect cell surface processes, but it will also have different flow characteristics that might reduce fragmentation or increase residence time in the fermenter. The mutant might derive some advantage from either or both of these two factors. The selective advantage need not be large because these cultures are maintained for more than 500 generations. In human terms, that's the equivalent of more than 12,000 years! Twelve thousand years ago we were just inventing farming.

The selective advantage of colonial mutants may be due to a combination of environmental factors related to the nature of the suspending medium and the flow characteristics of a continuous flow fermenter. A general genetic feature of long-term cultures of fungi like these, however, is the periodic appearance of new mutant populations. It happens in the Quorn™ fungus, *Fusarium*, and it has also been demonstrated in chemostat cultures of *Aspergillus niger* and *A. nidulans*. In all three fungi, increases in the mutant population occurred at intervals of approximately 30 generations. The mechanism that generates these periodic new mutant populations is not known, but it does have a compelling logic. The periodically increased mutant populations represent adaptive peaks in which a fungus that has been growing vegetatively in a uniform environment for a long time can introduce a burst of genetic heterogeneity into its population structure: a built-in contravention of Hardy-Weinberg rules!

9.10 Mycotechnology

Although the Quorn™ fermentation is the only liquid fermentation that yields fungal biomass as the sole food product, there are several other liquid batch cultures used to produce fungal metabolites. Antibiotic production by fungal fermentation remains an important commercial process, although it has been overshadowed in value by fungal metabolites with other pharmaceutical uses, like the cholesterol-lowering statins (total annual sales value about $5 billion) and immmunomodulatory drugs.

Fungal applied genetics is, obviously, aimed at improving production in industry; however, the production process must be fully understood if the right choice is to be made of the character to select to achieve improved production. Commercial enterprises jealously guard their own know-how, and such applied genetics tends to be done behind closed doors. This makes it difficult to find specific examples in the literature of how currently used strains were derived.

The oldest stories, which are still worth telling, relate to yeast alcoholic fermentations and penicillin production. Although *Saccharomyces cere-*

visiae yeasts are used in several industries it is worth remembering that the aims of improvement programs are quite different in different industries. Desirable characteristics for baker's yeast are rapid production of CO_2, so that the dough rises promptly, good flavor and aroma, low alcohol tolerance (up to 5%), but good temperature tolerance. Good bread yeasts can grow up to 42°C during leavening and ferment up to 55°C, into the early stages of baking. Wine yeasts should be able to produce 8–15% alcohol and therefore be alcohol-tolerant. They must also tolerate acid pH (3–4), and be resistant to the SO_2 used to control bacteria and wild yeasts. An ability to ferment well at low temperatures (13–15°C), to flocculate (so that the yeast cells settle out well at the end of fermentation), and to produce good flavor and bouquet chemicals without off-flavors are all essential characteristics of wine yeasts. *S. cerevisiae* strains are also used as top-fermenting beer yeasts, whereas lagers use the bottom-fermenting *S. carlsbergensis*. Beer yeasts work at around pH 5, and do not need high alcohol tolerance, but they must be resistant to hop oils and have a high rate of fermentation and good flocculation after fermentation. Beer yeasts are selected for flavor and aroma characteristics. Different yeasts are needed for the different types of beer and, indeed, to suit the different market preferences of the consumers.

The improvements made in penicillin production make an interesting story because we know it started in 1928 with Sir Alexander Fleming's first observation of the effect of penicillin on Gram-positive bacteria. The industrial fungus is *Penicillium chrysogenum*, and Fleming's original strain was able to produce 2 U (units) of penicillin per milliliter of culture. Natural isolates give penicillin yields varying from zero to more than 100 U/ml in surface culture and from 1 to 80 in submerged culture. Modern strains can produce 6000 U/ml! The ancestor of all industrial strains is a single haploid uninucleate conidium from a culture collected from a rotten melon. It gave yields of 100–200 U/ml when isolated in 1951. Selection can only work if heritable variation exists in the population. During the penicillin yield improvement program, genetic variability was produced by use of some of the most potent mutagens known: X-rays, ultraviolet light, nitrogen mustard, nitrosoguanidine, diepoxybutane, and others. Following mutagenesis, selection was made for improved adaptation to culture conditions, and for production of penicillins with different molecular structures. A useful "loss of ability" mutation that occurred during the program was an ultraviolet light-induced mutation unable to make the yellow pigment chrysogenin. This meant there was no longer any need to remove the pigment during purification and significantly reduced downstream processing costs.

Since the introduction of penicillin, many millions of chemicals and metabolites have been analyzed for antimicrobial and other pharmaceutical activities. Chemicals screening is a major activity of the pharmaceutical and agrochemical industries around the world. Antibiotics obtained from fungi that are presently of clinical use as antibacterial agents include the

still-important penicillin, cephalosporin, and fusidic acid (both of the latter are useful against penicillin-resistant bacteria), and the antifungal griseofulvin (used to control fungal infections of the skin, nails and hair). Antibiotics are obviously the products that come to mind first when thinking of medically useful fungal products. Several of the mushrooms cultivated in Asia, especially shiitake (the Japanese common name) or shiang-gu (the Chinese name of the fungus known scientifically as *Lentinula*) have been shown to produce materials with antitumor, anticancer, antiviral, antihypertensive, and anticholesterol effects.

Filamentous fungi in fermentation cultures are also used to produce many commercial enzymes, most of which are for use in processing foods. For example, amylases (from *A. niger, A. foetidus, Rhizopus foetidus*) are used to convert starchy substrates to sugars prior to alcoholic fermentation, and to make chocolate syrups from cocoa. Invertase (from *A. oryzae, A. niger*) is used for sucrose conversion in confectionery. Some other industrial uses for fungal enzymes include proteinases of *Aspergillus* and *Mucor* are used in cheese production, proteinases and lipases from *A. oryzae* used are in detergents, and proteinases are used in hide processing in the tanning industry. The global value of enzymes from filamentous fermentations is in the region of $1 billion per annum.

Apart from alcohol, the single most important fungal fermentation metabolite is citric acid, 300,000 tons of which are produced each year (mostly by *Aspergillus niger* and *A. wentii*) for use mainly in effervescent soft drinks. In addition, several human food products made from growing filamentous fungi on water-soaked seeds (i.e., soy sauce and various other fermented foods) are produced in Asia. In the fine-chemicals category, several pigments produced by *Monascus* are used as food colorants, and some fungi are used to perform specific chemical transformations on precursors of pharmaceutical products.

Mushroom crops are produced using solid-state fermentations that use compost, straw, or timber as the major substrate. Total world production of mushrooms in 1996 was estimated to be around 5.8 million tons, with China contributing 3.5 million tons to that total. *Agaricus bisporus* is by far the most commonly cultivated mushroom around the world. The *Agaricus* crop accounted for more than 70% of total global mushroom production in the mid-1970s, but today it accounts for something closer to 30% even though production tonnage has more than doubled in the intervening years. The biggest change during the last quarter of the twentieth century has been the increasing interest shown in a wider variety of mushrooms. So-called exotic mushrooms have appeared even in the most conservative of markets (e.g., the United Kingdom), and fresh *Lentinula* and *Pleurotus* (oyster mushroom) are routinely on sale alongside *Agaricus* in local supermarkets. Most of these mushrooms are cultivated fairly close to the point of sale. *Ganoderma lucidum* is unique among cultivated mushrooms in being consumed for its pharmaceutical value (real or imagined) rather than as a food. Under

the names *lingzhi* or *reishi*, several members the *G. lucidum* complex provide various commercial brands of health drinks, powders, tablets, capsules, and diet supplements. This is an important market because *Ganoderma* is highly regarded as a traditional herbal medicine in China, and its popularity has spread to other Asian countries, as well as to the wider world.

In all of these industries, genetic improvement of the fungal strain on which the industry depends has been key to improvements in yield, and the reliability and quality of desirable products, decreases in the formation of unwanted products, reduction in the costs of nutrients, equipment, fuel, labor, transport, and improvements in downstream processing. Genetic methods that are available will depend on the fungus, but they may include selection, mutation (spontaneous or induced), recombination (sexual or parasexual), heterokaryosis, and/or chromosomal and karyotype changes to allow complementation, epistasis, or hybrid vigor to be used in aneuploids, diploids or polyploids.

Today, some form of direct genetic modification is likely to be the first thought in any campaign aimed at strain improvement, but genetic modification has also opened up possibilities for entirely new processes whereby a fungal fermentation, which is a well-understood biotechnology, might be used to produce a commodity that originated in a different organism: a heterologous protein. In particular, brewing and enzyme-producing companies around the world possess highly sophisticated "in-house" fermentation technologies, and are investing in developing heterologous gene expression in fungi such as *Saccharomyces cerevisiae* and several species of *Aspergillus*. At the moment, emphasis is concentrated on high-value therapeutic proteins such as factor VIII, interferons, human serum proteins, and recombinant vaccines.

Among the attractions of these fungi as "protein-production machines" are that they can harbor multicopy plasmids, which confer a high gene dosage effect. Further, several well-characterized strong promoters, which efficiently drive transcription are available, and, importantly, these include both constitutive and regulated promoters. Promoters of the glycolytic pathway genes phosphoglycerokinase (PGK), glucose-6-phosphate dehydrogenase, and enolase are all constitutive. Regulated promoters are potentially more useful because their activity can be controlled during the fermentation process. Examples of these include the promoters for alcohol dehydrogenase, which is activated by glucose starvation, acid phosphatase, which is activated by phosphate limitation, yeast GAL1 and GAL10 promoters, which are activated by galactose in the growth medium, the copper thionein promoter, which responds to copper ions in the medium, and the heat-shock promoters, which are activated by a shift to high temperatures.

Several very effective transcription terminators are equally available (e.g., the 3′-end of the PGK gene), which are important because they limit transcription read-through into sequences in the vector-construct adjacent to the target gene. Most important of all is the ability to secrete foreign

proteins efficiently because this avoids intracellular proteolytic degradation, aids proper folding, and simplifies downstream purification of the final product. The desired route of secretion can be assured by fusing the coding region of the foreign gene with the appropriate fungal signal sequence so that a fusion protein is secreted. Downstream processing will then include removal of the signal sequence by proteolytic cleavage. The main deficiency of all fungi used for biotechnology at the moment is their inability to perform posttranslational modifications of proteins, including glycosylation, phosphorylation, or addition of fatty acid or farnesyl residues, which are sometimes required if a therapeutic protein is to be recognized as authentic in the patient's body. Insect cell lines and mammalian hosts, like genetically modified farm animals, are currently required for the production of such proteins.

Publications and Websites Worth a Visit

Anderson, J.B. & Kohn, L.M. (1998). Genotyping, gene genealogies and genomics bring fungal population genetics above ground. *Trends in Ecology and Evolution* **13**, 444–449.

Aravind, L. & Subramanian, G. (1999). Origin of multicellular eukaryotes—insights from proteome comparisons. *Current Opinion in Genetics and Development* **9**, 688–694.

Arnau, J. & Oliver, R.P. (1996). Genetic analysis in the imperfect fungus *Cladosporium fulvum*: the development of a map-based strategy for the cloning of genes involved in pathogenicity. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 349–361. Marcel Dekker, Inc.: New York.

Bennett, J.W. (1998). Mycotechnology: the role of fungi in biotechnology. *Journal of Biotechnology* **66**, 101–107.

Berbee, M.L. & Taylor, J.W. (1993). Dating the evolutionary radiations of the true fungi. *Canadian Journal of Botany* **71**, 1114–1127.

Berbee, M.L. & Taylor, J.W. (1999). Fungal phylogeny. In *Molecular Fungal Biology* (R.P. Oliver & M. Schweizer, eds.), pp. 21–77. Cambridge University Press: Cambridge, U.K.

Crute, I.R., Holub, E.B. & Burdon, J.J. (1997). *The Gene-for-Gene Relationship in Plant-Parasite Interactions*. CAB International: Wallingford, U.K.

Eslava, A.P. & Alvarez, M.I. (1996). Genetics of *Phycomyces*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 385–406. Marcel Dekker, Inc.: New York.

Farman, M.L. & Leong, S.A. (1996). Genetic analysis and mapping of the avirulence genes in *Magnaporthe grisea*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 295–315. Marcel Dekker, Inc.: New York.

Frank, S.A. (1992). Models of plant-pathogen coevolution. *Trends in Genetics* **8**, 213–219.

Frisvad, J.C., Bridge, P.D. & Arora, D.K. (1998). *Chemical Fungal Taxonomy.* Marcel Dekker, Inc.: New York.

Hibbett, D.S., Pine, E.M., Langer, E., Langer, G. & Donoghue, M.J. (1997). Evolution of gilled mushrooms and puffballs inferred from ribosomal DNA sequences. *Proceedings of the National Academy of Sciences of the U.S.A.* **94**, 12002–12006.

Hillis, D.M., Moritz, C. & Mable, B.K. (1996). *Molecular Systematics*. Sinauer Associates, Inc.: Sunderland, MA.

Hoekstra, R.F. (1996). Horizontal transmission in fungal populations. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 337–348. Marcel Dekker, Inc.: New York.

Johnson, J. (1999). Phylogenetic relationships within *Lepiota sensu lato* based on morphological and molecular data. *Mycologia* **91**, 443–458.

Lumbsch, H.T. (2000). Phylogeny of filamentous ascomycetes. *Naturwissenschaften* **87**, 335–342.

Lutzoni, F. & Vilgalys, R. (1995). Integration of morphological and molecular data sets in estimating fungal phylogenies. *Canadian Journal of Botany* **73**, S649–S659.

Milgroom, M.G. (1996). Recombination and the multilocus structure of fungal populations. *Annual Review of Phytopathology* **34**, 457–477.

Moore, D. & Chiu, S.W. (2001). Fungal Products as Food. In *Bio-Exploitation of Filamentous Fungi* (S.B. Pointing & K.D. Hyde, ed.), pp. 223–251. Fungal Diversity Press: Hong Kong.

Odds, F.C., Gow, N.A.R. & Brown, A.J.P. (2001). Fungal virulence studies come of age. *Genome Biology* **2**, reviews1009.1–1009.4. The electronic version of this article can be found online at http://genomebiology.com/2001/2/3/reviews/1009.

Page, R.D.M. & Holmes, E.C. (1998). *Molecular Evolution. A Phylogenetic Approach*. Blackwell Science Ltd.: Oxford, U.K.

Philippe, H., Germot, A. & Moreira, D. (2000). The new phylogeny of eukaryotes. *Current Opinion in Genetics & Development* **10**, 596–601.

Piercey-Normore, M.D. & Egger, K.N. (2001). Identification and genotypic characterization of filamentous fungi. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 1–22. Oxford University Press: Oxford, U.K.

Shaw, D.S. (1996). Genetic analysis in the oomycetous fungus: *Phytophthora infestans*. In *Fungal Genetics: Principles and Practice* (C.J. Bos, ed.), pp. 407–417. Marcel Dekker, Inc.: New York.

Sugiyama, J. (1998). Relatedness, phylogeny, and evolution of the fungi. *Mycoscience* **39**, 487–511.

Swift, R.J., Craig, S.H., Wiebe, M.G., Robson, G.D. & Trinci, A.P.J. (2000). Evolution of *Aspergillus niger* and *A. nidulans* in glucose-limited chemostat cultures, as indicated by oscillations in the frequency of cycloheximide resistant and morphological mutants. *Mycological Research* **104**, 333–337.

Taylor, J.W., Bowman, B., Berbee, M.L. & White, T.J. (1993). Fungal model organisms: phylogenetics of *Saccharomyces*, *Aspergillus* and *Neurospora*. *Systematic Biology* **42**, 440–457.

Tehler, A., Farris, J.A., Lipscomb, D.L. & Källersjö, M. (2000). Phylogenetic analyses of the fungi based on large rDNA data sets. *Mycologia* **92**, 459–474.

Thon, M.R. & Royse, D.J. (1999). Partial β-tubulin gene sequences for evolutionary studies in the Basidiomycotina. *Mycologia* **91**, 468–474.

Veuthey, A.L. & Bittar, G. (1998). Phylogenetic relationships of Fungi, Plantae and Animalia inferred from homologous comparison of ribosomal proteins. *Journal of Molecular Evolution* **47**, 81–92.

Walton, J.D. (2000). Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. *Fungal Genetics and Biology* **30**, 167–171.

Whelan, S., Li, P. & Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends in Genetics* **17**, 262–272.

Wiebe, M.G., Robson, G.D., Oliver, S.G. & Trinci, A.P.J. (1995). Evolution of *Fusarium graminearum* A3/5 grown in a series of glucose-limited chemostat cultures at a high dilution rate. *Mycological Research* **99**, 173–178.

Wolf, Y.I., Kondrashov, A.S. & Koonin, E.V. (2000). Interkingdom gene fusions. *Genome Biology* **1**, research0013.1–0013.13. The electronic version of this article can be found online at http://genomebiology.com/2000/1/6/research/0013

Historical Publications Worth Knowing About

Dawkins, R. (1989). The Selfish Gene, second edition. Oxford University Press: Oxford, U.K.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Sogin, M.L. (1989). Evolution of eukaryotic microorganisms and their small subunit ribosomal RNAs. *American Zoologist* **29**, 487–499.

CHAPTER 10

The Genetics of Fungal Differentiation and Morphogenesis

Revision Concepts

- The fundamental growth pattern of the Kingdom Fungi is polarized, invasive, and localized at the hyphal apex.
- For cell patterning, hyphal branching in fungi is the equivalent of cell division in plants and cell migration in animals.
- The origin of the branch and the direction of its growth determine the pattern of formation of fungal multicellular structures.
- Regulated gene expression is responsible for the transition routines between different cell types during development, differentiation, and tissue morphogenesis.
- Gene regulation can be imposed at any time during the process between DNA transcription and protein activity.
- Transient changes in genome activity are carried out during gene transcription by DNA-binding proteins in combination with regulatory DNA sequences.
- Transcription factors are characterized by peptide motifs (e.g., zinc fingers, helix–loop–helix and helix–turn–helix structures, and leucine zippers).
- The chromatin structure of DNA, based on the nucleosome subunit, is designed to maintain transcription at minimal levels in eukaryotes.
- Nucleosomes have a structural as well as positional influence on the level of gene transcription; SWI/SNF and ISWI are multiprotein switches that remove the nucleosome barrier to gene expression.
- Mediator is a multiprotein factor that interacts with RNA polymerase II and links initial transcription of protein-encoding genes with transcription elongation.
- Long-term changes in genome activity are carried out by enhancers, scaffold-attachment regions (SARs), matrix-associated regions (MARs), locus control regions (LCRs), and silent information regulators (SIRs).
- Silencing repeated chromosomal regions can be achieved by DNA methylation (which is uncommon in fungi), via transposable elements,

through quelling, or transvection (gene function altered by homologous chromosome pairing).
- Spliceosomes carry out posttranscriptional RNA modifications.
- Translational control is governed by the half-life of mRNA, the length of their poly-A tail, dephosphorylation of translation initiation factors, and untranslated regions (UTRs).
- Substrate, coenzyme, or ligand binding and covalent reactions are post-translational modifications that affect protein function.
- Ubiquitin functions in two ways as a marker for proteins destined for proteolysis.
- Protein interaction maps can be deduced with the use of the yeast two-hybrid system.
- A relatively small proportion of the fungal genome is involved in a particular morphogenetic process.
- Dimorphism is a morphogenetic process in which yeast, which normally grow as individual cells, adopt a hyphal growth pattern.
- The causality of dimorphism is hotly debated because it can be brought about by specific gene expression as well as by fairly nonspecific influences (e.g., temperature and nutrient shifts).
- Different fungi use very similar signaling pathways to arrive at very different responses to external stimuli.
- The majority of loci involved in conidiation are responsible for attaining competence for sporulation, but they are otherwise not specifically involved in conidiation.
- *Bristle* and *abacus* are two of the very few loci that are involved in distinct events in conidial morphogenesis in *Aspergillus nidulans*.
- Feedback fixation is the manner in which a regulatory sequence reinforces expression of the whole developmental pathway, making it independent of the external cues that initiate it.
- Ascomycete protoperithecial mutants suggest that achievement of a quiescent state is prerequisite to fruit body development.
- The relationship between monokaryotic and dikaryotic fruiting in basidiomycetes is not straightforward.
- In some basidiomycetes a portion of the developmental pathway for multicellular vegetative structures and sexual fruit bodies is shared, but the meiosis and spore formation pathways are entirely separate.
- Glucanases have an important role in fruiting body formation; other proteins, including hydrophobins, laccases, and phenoloxidases, are correlated with both asexual and sexual multicellular structures.
- Control mechanisms for fungal morphogenesis include mating-type factors, and may include translational triggering and feedback fixation.
- Normal morphogenesis in fungi may be a collection of distinct developmental subroutines, which can be put into operation independently of one another.

- Key stages for control of fungal morphogenesis are competence, induction, and change, but the overriding principle is tolerance of imprecision, enabling progeny spores to be produced even in the most stressful conditions.
- Gene regulation mechanisms may be specific to fungi at the phylum level.

10.1 Differentiation and Morphogenesis

Growth of the vegetative fungal hypha, showing polarized, invasive extension growth localized at the hyphal apex, is the fundamental growth pattern of all members of Kingdom Fungi, and of some members of related groups. If we can borrow a word used in everyday computer terminology, the fungal hypha is the *default* growth condition of the fungal genome. Vegetative hyphal growth requires coordinated expression of the components of the genome so that the whole of the growth process can be supported, located, and projected into the extension of the hyphal tip. All of this requires regulation of gene expression. Most fungi also produce a range of cell types that differ in cell shape and growth pattern. These require further programs in which gene expression is integrated into developmental routines involving transmission and receipt of signals to organize transitions between different cell types. Some of those signals will be intracellular, and some will be extracellular relating the nutritional and physical state of the environment, but all will require signal transduction pathways comprising receptor, transmission, and amplification and effector components.

This aspect of development, which is cell differentiation, depends on differential management of hyphal functions, part of which relies on genetic regulation leading to synthesis of gene products specific to certain cell types, but part of which can also include epigenetic phenomena including gene silencing as well as phenotype changes in which physical forces establish morphological change by altering cytoskeletal organization, for example. Such regulatory events are sufficiently robust to account for most hyphal differentiation including even that of such yeastlike fungi as *Saccharomyces cerevisiae*. The yeastlike cells can be interpreted as hyphal cells trapped in a highly differentiated yeast-form morphology in which the normal invasive hyphal apex growth is adapted to the pattern of growth recognized as budding. Even yeasts, however, can be induced to grow as elongated filaments, dedifferentiating to the default fungal invasive growth form.

Beyond cell differentiation, but obviously dependent upon it, we place fungal tissue morphogenesis. Even the vegetative fungal mycelium may be considered as a tissue because it grows outward into new territory and consequently has controlling signals that ensure that hyphae normally grow away from one another to form the typical "colony" with an outwardly migrating growing front. Tissue development requires that different hyphae cooperate in an organized way. For tissues to be formed the invasive

outward growth pattern of the vegetative mycelium must be modified so that independent hyphal apices grow toward each other, allowing their hyphae to branch and differentiate in a cooperative fashion. The structures to which the tissues contribute (e.g., spore-forming fruiting bodies) actually arise on the vegetative mycelium, so these changes in growth pattern must be localized, and they must be a response to regulatory processes that are imposed upon the vegetative mycelium.

Another aspect is that tissue formation demands that the continuous tube of hypha produced by the growing apex is divided up into cells or compartments by the formation of cross-walls (septa). This enables differentiation to be localized, offering the possibility that adjacent compartments might follow different pathways of differentiation, and even be of different size. Lower fungi (e.g., Zygomycotina like *Mucor*) have coenocytic hyphae. Although they do not form multicellular structures they do form septa at certain stages during development: the gametangia that eventually fuse and develop into a zygosporangium are separated from the rest of the coenocytic hyphae by septa, so that the zygosporangium develops alongside vegetative hyphae.

Fungi that do exhibit complex developmental pathways form septa at regular intervals in mycelial hyphae, but the septa usually have a pore (more or less central), which may be elaborated with the parenthesome apparatus in basidiomycetes, or are associated with Woronin bodies in ascomycetes. Although the septal pore is common feature, it is clearly the case that the movement or migration of cytoplasmic components between adjacent cells is under very effective control. There are instances in which nuclei move freely, but mitochondria do not, and others in which rapid migration of vacuoles is not accompanied by migration of any other organelle. Some biochemical experiments have even demonstrated that different sugars can be translocated in opposite directions in a hypha at the same time. There are also numerous examples available where grossly different pathways of differentiation have been followed on the two sides of what appear (to the electron microscope) to be open septal pores. Whatever the appearance of the open septa, the hypha can clearly be separated into cells whose interactions are carefully regulated and which can exhibit contrasting patterns of differentiation.

The hyphae of Ascomycotina and Basidiomycotina are characteristically divided up into cells by these septa-with-pores, but please don't forget that every fungal cell *is* just a segment of a tubular hypha. This is *very* important because the hyphal growth form must influence the characteristics of the controls that regulate fungal tissues. Filamentous hyphal growth can be interpreted on the basis of a regular cell cycle. By increasing the number of growing points, hyphal branching is the equivalent of cell division in animals and plants. Although plant morphogenesis depends on placement of the cross-wall, cross-walls are formed at right angles to the long axis of the hypha in fungal hyphae. Except in cases of injury or in hyphal tips already

differentiated to form sporulating structures, hyphal tip cells are not subdivided by oblique cross-walls, nor by longitudinally oriented ones. Even in fission yeast cells forced to produce irregular septation patterns under experimental manipulation, the plane of the septum is always perpendicular to the plane including the longest axis of the cell. In general, then, the characteristic fungal response to the need to convert the one-dimensional hypha into a two-dimensional plate or three-dimensional block cannot depend on a different geometrical arrangement of the septum. The only solution open to the fungal hypha is the formation of branches. The septum in the branch will still be formed at right angles to the long axis of the branch, but its orientation relative to the parent hypha will depend entirely on the positioning of the branch apex, which is established some time prior to septum formation.

As a result, there are two fundamental processes involved in construction of fungal multicellular structures: the first is the origin of the branch (i.e., its appropriate placement and orientation on the parent hypha) and the second is the direction of growth of the new hyphal apex that is created by the branching event. The former process seems to be the formal equivalent of determination of morphogenetic growth by orienting the plane of division and the new cross-wall as is seen in plants, and the latter has much in common with the morphogenetic cell migrations that contribute to development of body form and structure in animals. Viewed in this light, therefore, the fungal Kingdom is seen as employing morphogenetic processes that have affinities with both of the other major eukaryote Kingdoms. There is no substantial difference in the nature of the questions that need to be answered in studies of development in the three eukaryote Kingdoms. How do genes act to establish basic cell behavior? How do cells become different? How do cells influence one another? How do cells cooperate to form structures? An animal embryologist asks the same questions as a developmental mycologist. The answers may be different in detail, with the details being determined by the lifestyle according to Kingdom-specific adaptations of the organism concerned. There is likely to be an underlying similarity in strategy, however, because the same basic eukaryotic cell structure is used throughout, and in eukaryotes most gene regulation occurs at the initiation of transcription.

10.2 Genetic Approaches for Analyzing Gene Regulation

Gene regulation can be imposed at any of the stages in the flow of information from the DNA to the working protein: by controlling which genes are transcribed into RNA, by regulating which RNA products are spliced to make functional messenger RNA, by determining which mRNAs are transported to the cytoplasm, by regulating mRNA translation into protein, and then by regulating the function and lifetime of the protein itself. There

is, indeed, evidence for gene regulation at each of these stages, but despite examples of controls at other levels, there is a great deal of evidence to show that transcriptional control is the most critical and widely used level of gene regulation in eukaryotes.

Analysis of regulatory factors focuses on mutations that affect gene function without affecting the primary structure (amino acid sequence) of the gene product. A gene responsible for a phenotype that is sensitive to the amount of gene product produced in the cell is the best candidate. After choosing the "target gene," the experimenter searches for mutations that affect expression of that gene; these are regulatory mutations. Regulatory mutants that map within the target gene, or in the immediate vicinity it, can indicate DNA sequences that influence transcription. Such DNA sequences may serve as attachment sites for DNA-binding proteins that regulate transcription. They are called *cis*-acting elements because they work on the same DNA molecule as the target gene. Promoters are *cis*-acting elements to which the RNA polymerase binds. Another type of *cis*-acting element is the *enhancer*, which is a binding site for proteins that control the level of transcription. Sequences like this can be studied in the laboratory by using reporter constructs. These replace the target gene with the coding region of a heterologous gene that produces an easily identifiable product (the so-called reporter).

The β-galactosidase gene from the *Escherichia coli lac* operon (which can be detected through its reaction with chromogenic substrates: a colorless substrate known as X-Gal is turned blue in the presence of β-galactosidase) are popular reporters because of their colored products, as is the green fluorescent protein (GFP) isolated from the luminescent jellyfish *Aequorea victoria*, which absorbs blue light and re-emits it as green fluorescence. Color variants have been prepared that provide the opportunity for dual-labeling studies. There is a red fluorescent protein available that is isolated from *Discosoma* spp., an IndoPacific sea anemone. Along with the reporter gene, the whole reporter construct will include the regulatory regions of the target gene so that when the construct is reintroduced into the target genome by transformation, the effect of *in vivo* regulatory factors can be tested. Systematic mutagenesis (called *site-directed mutagenesis*) across the presumed regulatory region can then be used to study the influence of each base pair in the regulatory sequence.

Regulatory genes located on a different DNA molecule to the target gene are *trans*-acting elements. They are structural genes for polypeptides, known as *trans*-acting factors, that interact with the *cis*-acting elements of the target gene. *Trans*-acting proteins that regulate transcription are generally known as *transcription factors*. Mutations in *trans*-acting elements will alter the level of target gene expression (or expression of a reporter construct), but genetic mapping will locate them away from the site of the target gene or reporter. With *in vitro* techniques it is possible to isolate these proteins that bind to the DNA sequence of the *cis*-acting element.

10.3 Regulating Gene Expression: DNA-Binding Proteins

The basic structure of a typical eukaryotic protein-coding gene (Fig. 10.1) includes several different components. The protein-coding regions may be in two or more exons separated by introns that are spliced out of the RNA transcript and are untranslated. Regulatory sequences, where gene-controlling transcription factors bind, are mainly just upstream (i.e., on the 5′-side) of the transcribed region, although there may be other control regions, lying far outside the gene, that play a role in regulating chromatin structure. A common theme in eukaryote gene regulation is the involvement of DNA-binding proteins, which are involved in all aspects, including deciding which of the genes are to be expressed and for synthesizing the RNA transcripts of genes that are expressed, and a very large proportion of which have been identified from molecular genetic analysis of *Saccharomyces cerevisiae*.

These proteins bind to specific sequences in DNA and then interact with other proteins to activate transcription (Fig. 10.2). They have two structural domains that enable them to do this: a DNA-binding domain and a transcription-activator domain. Some activators have a third domain that reacts

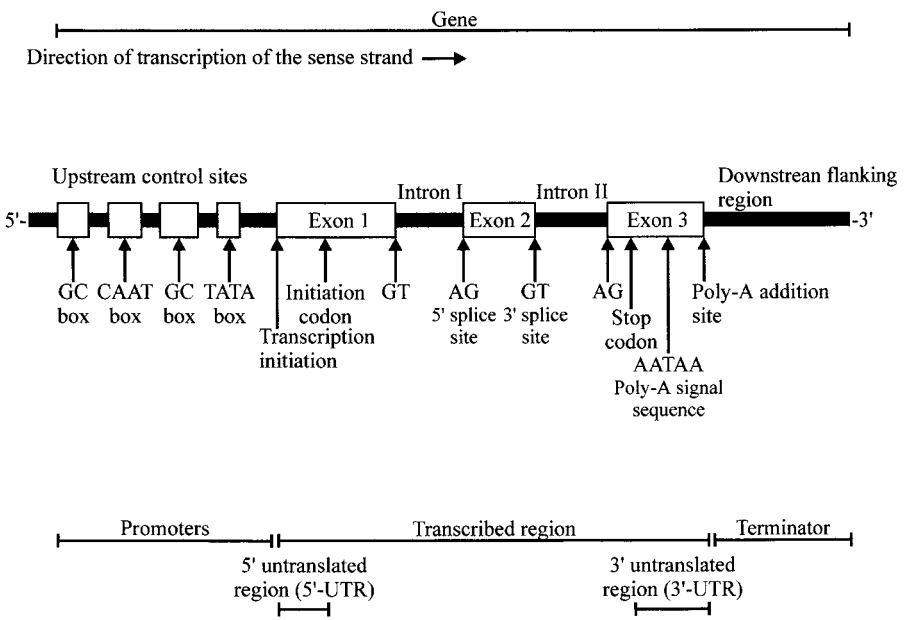


Fig. 10.1. The basic structure of a typical eukaryotic gene. The schematic diagram indicates the structure of a type II gene (i.e., a protein-encoding gene transcribed by polymerase II). The diagram is not drawn to scale and the relative sizes of the different sections differ between genes and between the eukaryotic Kingdoms.

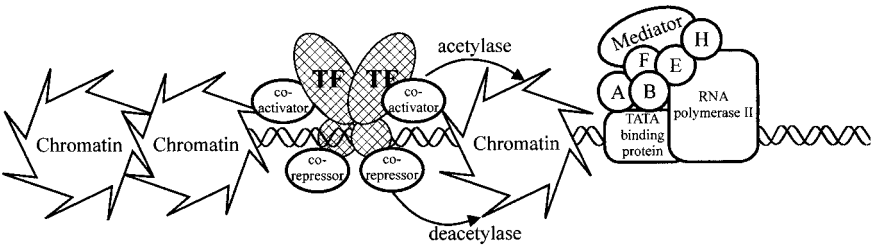


Fig. 10.2. Simplified illustration of the transcription machine. The polymerase, basal transcription factors (labeled A, B, E, F, and H), TATA-binding protein, and mediator (together with other proteins not represented here) occupy the core promoter of a gene. Upstream of this location histone octamers are represented by the tailed octagons labeled "chromatin." A transcription factor (TF) can interact with coactivators that recruit acetylase to acetylate the histone tails and open up the chromatin structure, or with corepressors that recruit a deacetylase, which restores chromatin structure.

to other specific signals (e.g., hormones or other signaling molecules). When such a molecule binds to these activators they cause an allosteric change that greatly increases the affinity of the protein for its DNA target sequence. This sort of control permits rapid changes in gene expression, enabling the cell to respond to external signals and transient changes in its metabolic circumstances. Rapid control over transcription factor activity of this sort often underlies the ability of extracellular conditions and signaling compound to control events going on within the cell. There may be an indirect activation when the extracellular signal interacts with a cell surface receptor that transduces the message to the cell interior, or a direct activation if the extracellular signaling molecule can enter the cell to interact immediately with a transcription factor or signal transducer.

DNA-binding domains in many different transcription factors share particular peptide motifs involved in the DNA helix binding function; these configurations are called the *zinc-finger* (in which an atom of zinc is conjugated to two cysteines and two histidines in the polypeptide), the *helix–loop–helix*, and the *helix–turn–helix* (which orient α-helixes of the polypeptide so that they can fit into the major groove of the DNA helix). Zinc-finger proteins generally have several "fingers," each of which is able to interact with a specific DNA sequence. There are also some common features in the transcription activation domains (e.g., being relatively rich in the amino acid asparagine or, alternatively, rich in proline). The shared features are associated with the general function of these molecules as transcription factors.

Other, much more subtle, aspects of their primary and secondary structures provide each one with its specificity for its DNA target sequence and the particular part of the transcription machinery it affects. Most activators

in eukaryotes must form dimers to function, and the functional proteins may be *homomers* (multimeric proteins composed of the same subunit) or *heteromers* (multimeric proteins composed of different polypeptide subunits). Heterodimerization increases the number of transcription factors that can be assembled from available monomers. Dimer formation depends on yet another characteristic domain of transcription factors (i.e., the dimerization domain), which is optimized for very specific interactions between particular polypeptides. The most common primary structure motif in dimerization domains is the leucine zipper. This is a sequence of amino acids that forms into an α-helix with leucine residues extending from the helix at regular intervals. The leucine zipper of one polypeptide can interlock with the leucine zipper of a second polypeptide, like the clothing version. Specificity for the "zipping" depends on the amino acids situated between the leucines.

In effect, there are two stages to transcription: transcriptional initiation and transcriptional elongation. Intrinsic to the initiation step are the specific interactions that determine which gene is expressed and which assemble all the proteins that will copy, or assist in copying, the gene into an RNA transcript. The second stage is the transcription process itself, during which the RNA polymerase translocates along the gene that produces the primary RNA transcript as a direct complementary copy of the gene.

10.4 Regulating Gene Expression: Chromatin Remodeling

One of the defining features of the eukaryotes is the possession of chromosomes, and the DNA packaging in the chromatin that makes up the chromosomes has an enormous influence on gene regulation in eukaryotes. Chromosomes in eukaryotes consist of about one-third genomic DNA, one-third histone proteins, and one-third nonhistone proteins. *Chromatin* is the name given to the complex between DNA and the proteins that make up the chromosome structure. An important function of chromatin is to reduce basal transcription of all genes to a very low level. In eukaryotes the normal structure of chromatin is entirely sufficient to maintain transcription at the minimal, basal level.

The basic structural unit of chromatin is the nucleosome, which consists of an octamer of histone proteins (two each of H2A, H2B, H3, and H4) around which is wrapped approximately 200 bp of DNA. Histone H1 binds to short stretches of DNA between nucleosomes, and helps maintain chromatin structure. Nucleosomes interact to construct further, higher, levels of chromatin fiber structure: from nucleosomes, to 10 nm fibers, then 30 nm fibers, on to chromosome loops, and ending with fully condensed metaphase chromosomes, which are the most compact form of DNA packaging in eukaryotes.

Chromosomes become less compact after completion of nuclear division, but there is a higher order folding (above the level of the 30 nm fiber) in interphase chromosomes. Heterochromatin is in a permanent state of compact folding. It is so compact, in fact, that proteins needed to activate gene expression cannot access the DNA. Constitutive heterochromatin is the DNA that contains no genes in centromeric and telomeric regions. Facultative heterochromatin is DNA that contains genes that are temporarily inactive because of the stage of development or position in the cell cycle. Regions of DNA that contain active genes are called *euchromatin*. Euchromatin consists of loops of 30 nm chromatin fibers that are equivalent to lengths of about 40–100 kb DNA. AT-rich DNA regions called MARs (matrix-associated regions) or SARs (scaffold-attachment regions) attach the loops to a protein network, called the *nuclear matrix*, that fills the nucleus.

Nucleosomes have an overriding influence on transcription because the DNA packaging within them represses gene expression. Transcription is made possible by specific positive regulatory mechanisms that rearrange nucleosome structure. Then, even when a specific gene is made accessible, the precise positioning of nucleosomes in the immediate vicinity influences transcription of it. This reflects a difference in regulatory strategy between prokaryotes and eukaryotes. Prokaryotes generally use negative regulation, effected by gene-specific repressors acting at structural gene promoters. Such a mechanism is arguably inadequate for the large genomes of eukaryotes because such a large number of different repressors would be needed to control gene expression. Instead, eukaryotes have adopted a mechanism that features general repression of the genome and requires integrated activation of transcription as the basis for cell-type–specific regulation. Nucleosomes repress transcription by covering protein-binding sites of DNA, thereby interfering with the interaction of the entire collection of DNA binding proteins, regulators, polymerases, and transcription factors required for transcription. Chains of nucleosomes can also become involved in higher-order coiling and thereby repress transcription of large chromosomal regions. Interactions between nucleosomes and other chromosomal proteins produce heterochromatin, in which gene expression is also repressed.

The molecular foundation of repression by nucleosomes lies in the configuration of the histone molecules, each of which has a characteristic "histone fold" and an N-terminal "tail." The histone folds keep the DNA in a central core particle, and it is this that prevents access of other DNA-binding proteins. The tail protrudes outside the core particle, taking part in the interactions that produce higher-order coiling. This is the basis of its involvement in gene activation. Acetyltransferase enzymes acetylate the histone tails, producing a chemical modification characteristic of transcribed chromatin. The acetyltransferases therefore serve as coactivators, stimulating transcription by lifting the repression caused when the core par-

ticles take on higher-order structure. Histone deacetylase enzymes do the reverse. They act as corepressors by removing the acetylation of the tail and thereby allowing the chromatin to take on the repressive higher-order structure (Fig. 10.2).

Histone acetylation, however, is not sufficient in itself for transcriptional activation because it does not disrupt the core particle of the nucleosome. Most inactive genes have their promoters occluded by nucleosomes. There are two multiprotein "chromatin remodeling complexes" that rearrange the structure of chromatin in an ATP-dependent manner to remove these promoter-blocking nucleosomes. One, known as "switch" and symbolized SWI/SNF, disturbs the core-particle structure; the other, "imitation switch" or ISWI, shifts the locations of nucleosomes on DNA. Nucleosome positioning is important in modulating gene expression. In yeast, the SWI/SNF complex is the first coactivator to arrive at a gene at which transcription is to be induced. The gene-specific activator proteins, which also recruit an acetyltransferase to acetylate the histone tails, recruit them, and the resultant loosening of the chromatin allows general transcription factors to get access to promoter regions. From that point transcription rapidly accelerates. The SWI/SNF complex is one of many that are involved in remodeling chromatin at specific chromosomal locations and in specific cells at particular points of development. Closely related protein complexes able to influence nucleosome position and/or structure are found in human and yeast cells, showing that chromatin remodeling machinery has been conserved throughout evolution.

10.5 Regulating Gene Expression: Transcription

Regulation of transcription in eukaryotes depends on multiprotein complexes assembled at DNA control sequences immediately adjacent to the start site of transcription, called the *promoter*. For many protein-coding genes the promoter contains the TATA box, which is a binding site for the constitutively expressed general transcription factor (GTF) called *transcription factor TFIID* (TF = transcription factor, II = for RNA polymerase II, D = TFII type D). Binding of TFIID to the promoter is critical to the assembly of a basal, stable transcriptional complex, which is able to recognize core promoter elements. This provides low levels of accurate transcription, called basal transcription using a variety of other transcription factors (TFIIA, B, E, F and H, see later) and RNA polymerase II itself.

These basal transcription machines are the globally used part of the transcription mechanism. Basal transcription is activated by a highly varied and very large group of transcription factors that assemble at distant enhancer sites. Such transcription activators provide the gene specificity and cell-type specificity of transcription. Even this transcription machinery, however, which might be composed of 40 or more polypeptides, is still dependent on

a third class of transcription factors called coactivators, which do not have site-specific DNA-binding ability by themselves, but act as intermediaries in the action of transcription activators on the basal transcription machinery (Fig. 10.2).

Following chromatin remodeling, the polymerase and accessory factors interact with the promoter (Fig. 10.2). There are three DNA-dependent RNA polymerases in eukaryotes, designated pol I, pol II and pol III, although we know most about pol II, which is responsible for all messenger RNA synthesis. Pol I transcribes the genes for ribosomal RNA and pol III transcribes transfer RNA (tRNA). RNA polymerases are complexes of 12 protein subunits, which require 23 other polypeptide transcription factors to recognize a promoter and *initiate* transcription. There is a perfect one-to-one correspondence between the components of the yeast and human systems, and components from animal cells function in yeast, which indicates a high degree of functional conservation of the transcription apparatus during eukaryotic evolution. These polypeptides, however, are insufficient to promote transcription *elongation*; an additional coactivator activity is required as an interface between activators and polymerase II, transducing regulatory information from enhancers to promoters. This factor is called "Mediator" and is a 20-subunit complex in yeast; corresponding complexes from mammals vary in subunit composition, but they are otherwise functionally the same.

There are several recognition sequences in the DNA, which are recognized either by the RNA polymerase itself or by a DNA-binding protein, which enable the transcription initiation complexes to be constructed at the correct positions on the DNA molecule. Bacterial RNA polymerases bind to promoter sequences, located immediately upstream of the gene to be transcribed. The "average" (or consensus) of all promoter sequences in *Escherichia coli* shows two six-nucleotide sequences; one is called the –35 box and has the sequence 5′-TTGACA, and the other is the –10 box, with the sequences 5′-TATAAT. The boxes are named for their position relative to the nucleotide base at which transcription begins, which is called +1. There is a stretch of 15 to 17 bases between the two boxes, which brings the two sequence motifs to the same face of the double helix, and ensures both can most effectively interact with the DNA-binding factor component of the RNA polymerase.

Eukaryotic promoters are more complex and there may be several sequences that are important in initiation of transcription of a gene. The 5′-TATA box is located about 60–120 base pairs upstream from the transcription start nucleotide in yeast (only about 30 base pairs in mammals), and this site directs the polymerase to begin transcribing. It is the binding site for the TATA-box binding subunit (TBP), plus more than eight *trans*-acting factors (TAFs), which together make up transcription factor II (TFII), which is one of the general transcription factors for RNA polymerase II. There are several of these, differing in function according to the

nature of their components, and they are distinguished by letter-suffixes. TFIID is responsible for promoter recognition. It uses the TBP subunit to bend DNA in the TATA-box region and enables interaction with TFIIB, which positions the polymerase on the promoter. TFIIH includes ATP-dependent helicases that unwind the promoter around the start site to trigger the initiation of transcription. They then maintain a "bubble" of unwound DNA around the nucleotide polymerization site, allowing pairing of the RNA product with the template through base pairing of about eight residues immediately adjacent to the polymerization site.

The mediator complex subsequently interacts with polymerase II to form the "holoenzyme" able to continue elongation. The switch from transcriptional initiation to elongation is associated with the phosphorylation of the carboxy-terminal tail of RNA polymerase II. Several elongation factors have been identified, and some may regulate transcription by interacting with sequences in the RNA transcript. Indeed, there is a very close connection between transcription and mRNA processing. The phosphorylated tail of RNA polymerase II in elongation mode interacts directly with factors involved in mRNA capping, 3′-end processing, and even splicing. By so doing, the various RNA-processing components are recruited to the transcription elongation apparatus and the RNA transcript, producing an "mRNA factory" in which synthesis and processing of mRNA are integrated.

About 100 base pairs upstream of the transcription start site a 5′-CCAAT box is also involved in promoter activity. A GC-rich sequence (consensus 5′-GGGCGG) about 100 base pairs further upstream may also serve as a promoter element. Further away, enhancer or activating sequences can be found. Enhancers are regulatory sites that can act at a distance and may be located many thousands of nucleotides away from the promoter. They may also be able to operate either upstream or downstream from the promoter they control. An enhancer is unable to drive transcription by itself, but it can enhance the activity of the promoter by several orders of magnitude. According to the nature of the transcription factor involved, this enhancement may occur in all cells (if that particular enhancer is bound by constitutively expressed transfer factors) or may occur only in a specific tissue or in response to a specific signal if the enhancer binding site is for factors that are involved in differentiation. In yeast, enhancer elements are usually called upstream activation sequences (UASs). Operation of enhancers can be tissue-specific and/or specific to environmental conditions, but they need to be intact to ensure maximal rates of transcription. They are the sites to which some other *trans*-acting (or transcription) factors bind to assist RNA polymerase to construct a preinitiation complex in a manner specific to a particular gene or gene family (e.g., the yeast *GCN*4 transcription factor, which, in response to amino acid starvation, activates transcription of many genes involved in amino acid synthesis by binding to a common UAS).

10.6 Galactose Utilization in Yeast: The Epitome of Eukaryote Regulation

Six coordinately regulated structural genes encode the proteins needed for hydrolysis and utilization of galactose and the galactose-containing disaccharide, melibiose, by *Saccharomyces cerevisiae*. Both sugars are highly relevant to the natural environment of yeast because they occur in plant exudates, especially in the nectaries. These genes are among the most tightly regulated genes known. Together with the fact that they control a shift in metabolism from one sugar to another, this accounts for the high level of interest in the system.

The GAL genes that controll utilization of galactose as a carbon source include the three structural genes that forming a tightly linked cluster, on chromosome II, *GAL*1, *GAL*7, and *GAL*10. *GAL*1 encodes the enzyme galactokinase, which phosphorylates galactose to galactose 1-phosphate (Fig. 10.3). *GAL*7 and *GAL*10 cooperate in the next step, converting galactose 1-phosphate to glucose 1-phosphate. GAL7 encodes galactose 1-phosphate uridyltransferase, which converts galactose 1-phosphate and UDP-glucose into glucose 1-phosphate and UDP-galactose. The *GAL*10 protein is an epimerase that regenerates UDP-glucose from UDP-galactose. All three enzymes are essential for metabolism of galactose. When galactose is present in the medium *each* of these proteins can represent up to 1.5% of the total soluble protein in the cytoplasm. Two other important components of the pathway, not physically part of the gene cluster, but certainly part of the regulatory circuit are *GAL*2, which encodes the galactose transporter, and *GAL*5, which is responsible for the phosphoglucomutase that performs the conversion of glucose 1-phosphate to glucose 6-phosphate. Glucose 6-phosphate can then directly enter glycolysis.

The *GAL* structural genes are tightly regulated at the level of transcription by whatever carbon sources are available. In fact, there is a dual control: enzyme synthesis is induced by galactose, but with an overriding repression by glucose. Glucose affects cell regulation in several ways, and may result in changed gene expression (called *glucose repression* or *catabolite repression*), but it may also affect mRNA turnover or directly influence individual enzymes (e.g., catabolite inactivation). Catabolite repression is the means through which glucose represses the expression of genes needed for utilization of alternative carbon sources. The regulatory protein, Mig1p, imposes glucose repression. Binding sites for Mig1p are located in the promoters of several glucose-repressed genes, including *GAL* genes. Even in the presence of galactose, glucose causes repression of *GAL* genes because growth with glucose causes Mig1p to repress expression of *GAL*4. The consequentially reduced levels of Gal4p result in low-level transcription of *GAL* genes. In addition, the Mig1p repressor inhibits function of any Gal4p

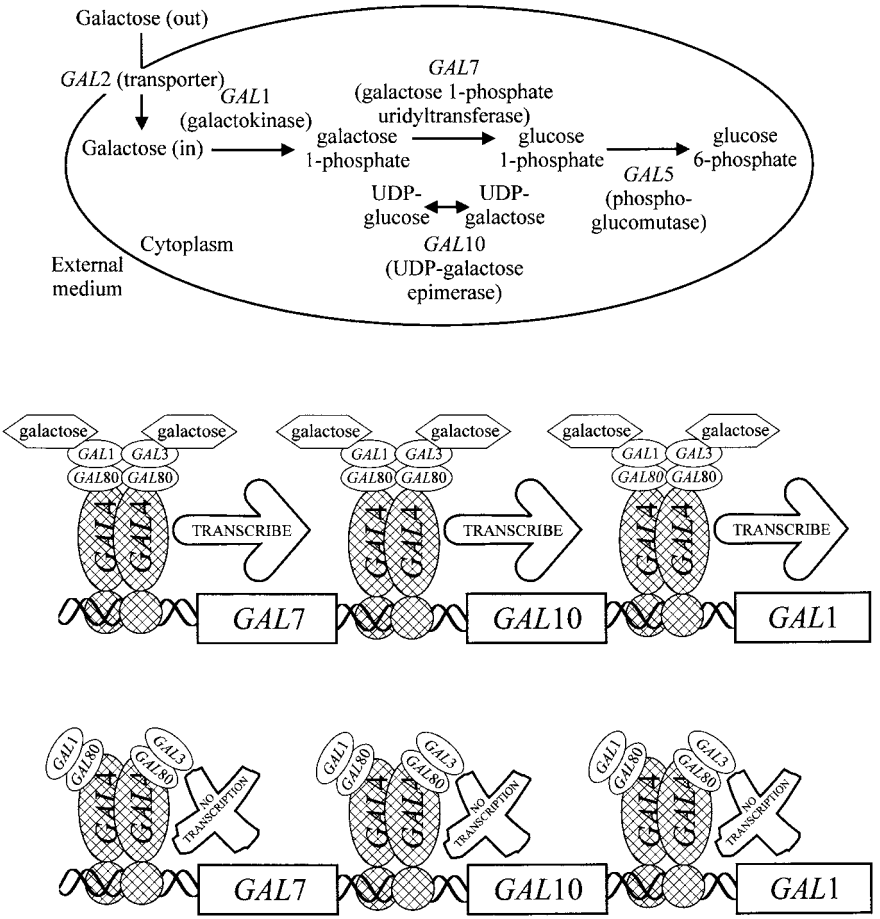


Fig. 10.3. Galactose metabolism and its regulation. The upper panel shows the biochemical reactions involved in uptake of galactose from the medium and its introduction into energy-yielding metabolism. The lower diagrams depict the role of the *GAL*4p transcription factor in transcription of the genes encoding the main galactose metabolizing enzymes and its interaction to the coactivators *GAL*1p and *GAL*3p and the co-repressor *GAL*80p.

that is produced by binding at an upstream repression sequence in the *GAL*1 promoter.

The *GAL*4-encoded protein (Gal4p) is a central component of the mechanism of *GAL* gene regulation (Fig. 10.3). Gal4p is a transcription factor, which binds to a UAS (enhancer) upstream of the promoters of the structural genes involved in galactose metabolism. The fact that the *three* structural genes are controlled by a *single* enhancer was indicated by mutations at a site located several hundred base pairs upstream of the gene cluster

affecting all three gene products. The *GAL*4 locus produces a *trans*-acting genetic regulator of *GAL*1, *GAL*7, and *GAL*10, which binds to the enhancer and coordinately activates transcription of all three genes. Mutations in *GAL*4 prevent its product from functioning as an activator, the phenotype being that none of the galactose-metabolizing enzymes are produced in the mutants when galactose is sole carbon source. Another *trans*-acting genetic element (called *GAL*80) was identified by the fact that *GAL*80 mutants continued to express all three galactose-metabolizing enzymes in the absence of galactose (constitutive expression). This indicated that the protein product of the *GAL*80 gene serves as a joint repressor of *GAL*1, *GAL*7, and *GAL*10. This integration of the control of several physically different genes by single control factors is what makes this galactose utilization pathway such a useful model for eukaryote genetic regulation.

The *GAL*80 gene product is not a DNA-binding protein, but it does bind to Gal4p; however, when Gal80p binds to Gal4p the latter loses activating ability. The crucial genetic test of this dependency between the two *trans*-acting elements was to construct double mutants that were defective in both regulatory genes. No induction of *GAL*7, *GAL*10, or *GAL*1 occurs in such *GAL*4$^-$, *GAL*80$^-$ double mutants. Remembering that *GAL*80$^-$ single mutants showed constitutive expression, the phenotype of the double mutant shows that *GAL*4 is epistatic to *GAL*80. The product of the *GAL*80 gene can evidently only function in the presence of the *GAL*4 protein.

One final component in the regulatory circuit is *GAL*3, which produces another polypeptide that is a coactivator of the structural genes of the enzymes that is only active in the presence of galactose. In normal yeast cells growing on glucose this regulatory system ensures that the galactose metabolic enzymes are not produced. This is because although both *GAL*4 and *GAL*80 *trans*-acting polypeptides are produced under these conditions, and the *GAL*4 product binds to the enhancer site, *GAL*80 protein binding to the *GAL*4 product quenches the activating ability of the latter. The cytoplasm also contains the *GAL*3 polypeptide, but this is inactive without galactose, and a very small amount of *GAL*1 product, which is produced through basal transcription. Remove the glucose and add galactose to this system and the whole picture changes. Galactose binds allosterically to the protein products of both *GAL*1 and *GAL*3, and the allosteric conformational change gives both polypeptides a high-affinity–binding site for the *GAL*80 protein. When these two polypeptides bind to the *GAL*80 product, the complex becomes dislodged from the activation domain of the Gal4p, and it is able to activate transcription of the *GAL*1, *GAL*7, and *GAL*10 gene cluster (Fig. 10.3). The galactose-metabolizing enzymes are synthesized and galactose can be used as the sole carbon source. There is an element of autocatalytic control in this mechanism because the *GAL*3 polypeptide is highly homologous to *GAL*1, but it lacks the galactokinase activity. Indeed, once *GAL*1 has been induced, the *GAL*3 protein is no longer essential.

Galactose eventually becomes depleted, and the last few galactose molecules detach from the *GAL*1 and *GAL*3 products. Without their allosteric effector, these molecules revert to the conformation in which they have no affinity for Gal80p, so the complex falls apart. This allows Gal80p to move back into the position that blocks the Gal4p activation domain; without the activation, transcription of the three *GAL* structural genes is reduced to basal, repressed levels (Fig. 10.3).

This example illustrates extremely well how eukaryotes use a complex of *cis*-acting and *trans*-acting elements, activators, repressors, coactivators, and allosteric interactions to achieve the relatively straightforward task of "switching on and switching off" the metabolism of the sugar galactose. There is an irony to this particular complexity. The appearance in yeast of the enzymes needed to metabolize galactose as a result of the addition of the substrate to the medium was the first instance of enzyme adaptation without cell division to be published, and that paper appeared in 1900. Working out the details took a further century. The system as described earlier is often called a *regulon*, a term that is used in imitation of the term *operon*, which describes prokaryote regulatory units. There is no real homology, however, between the galactose operon of *Escherichia coli* and the galactose regulon of *S. cerevisiae*; they operate on different principles and have different impacts on the cell. We think it is dangerous to use a vocabulary that might suggest that there could be some relationship. The most recent analyses of the yeast galactose utilization pathway using DNA microarrays and quantitative proteomics have shown that a thousand mRNAs change in response to changes in the galactose pathway. This indicates the complexity of the eukaryote regulatory and metabolic networks (Fig. 10.4).

10.7 Regulating Gene Expression: Repression and Silencing

Our emphasis so far has been on positive regulation (i.e., the activation of gene transcription), but gene expression must also be switched off and transcription factors can have negative effects as well as positive. A transcription factor that suppresses activation of transcription is called a *repressor*. Repressors may interact with general transcription factors (e.g., TFIIB and TFIID), thereby affecting assembly of the transcription complex. They may also interact with a corepressor, which recruits a histone deacetylase so that the promoter is silenced by re-establishment of the chromatin structure. Some repressors bind to the same DNA sites as activator proteins and cause repression by competing with activator proteins for binding, thereby preventing the activators stimulating transcription initiation. In these cases, the inhibitory factor acts by neutralizing the activity of a positively acting factor.

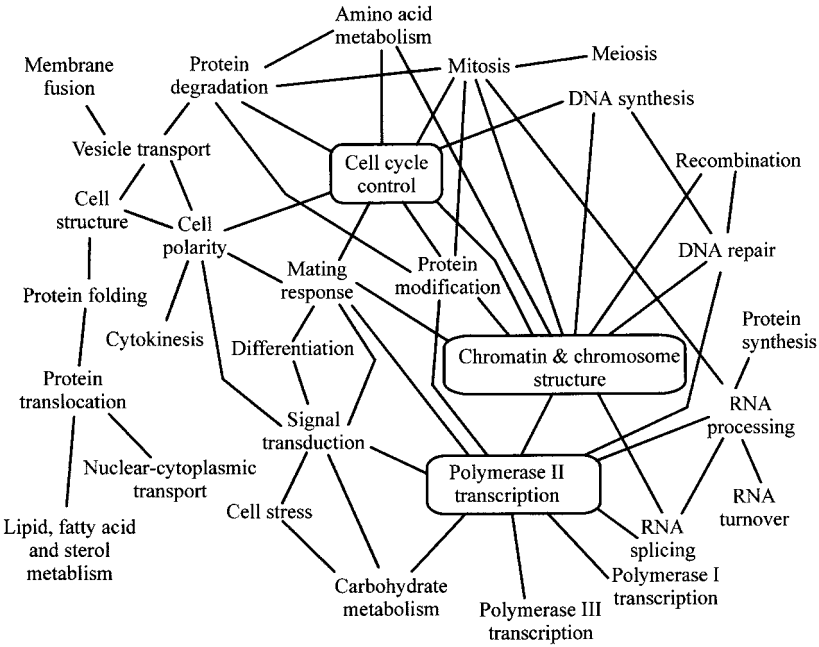


Fig. 10.4. A functional group interaction network for some groups of yeast proteins. Each connecting line represents at least 15 interactions between proteins of the connected groups. (Based on Fig. 2 in Tucker, Gera & Uetz (2001), Trends in Cell Biology **11**, 102–106.)

Nonetheless, negatively acting factors have an important role in transcriptional regulation. Silencers are *cis*-acting DNA sequences that, like enhancers, are the recognition sites for transcription factors. Silencers, however, are bound by repressors that inhibit activators and reduce transcription, inhibiting gene expression indirectly. As more examples have come to light, however, it has become evident that many transcription factors can act as either activator or repressor, depending on the gene being regulated and the cell type in which it is expressed. Indeed, in no case of transcriptional repression studied so far has the possibility been eliminated that repression results from neutralization of an activator. This has led to the tendency to drop the description "silencer" and call all *cis*-acting elements "enhancers," even though some may sometimes have repressors bound to them. The word *silencing* has taken on other meanings implying longer-term regulation (see later).

Transcription factors are polypeptide, and one way of reducing transcription of a gene or genes is to reduce the level of synthesis of the transcription factor. This does not provide rapid control over expression of the target genes because some time is required to downregulate the structural

gene for the transcription factor and reduce its concentration in the cell. As a result, this type of control tends to be associated with transcription factors responsible for the longer-term patterns of gene expression related to differentiation and morphogenesis. The specificity of already-synthesized transcription factors can be altered much more rapidly if they are sensitive to modification by other molecules in the cell. We have already referred to an example of this in relation to the yeast mating type α-2 polypeptide, which, in haploid α-cells, acts as a repressor of a-determining genes by binding to their enhancers; however, in diploid cells, the same α-2 protein dimerizes with the a-2 gene product to form a repressor of haploid-specific genes.

Some repressors function without binding to DNA; rather, they bind to a specific transcription activator and change its function. The process is called *quenching*, and the regulator protein may quench the DNA-binding activity of an activator (preventing attachment to the enhancer), or may block the activation domain of the activator. We have discussed an example of the latter type of repression in the yeast *GAL* system, where Gal80p binding to Gal4p quenches activation by Gal4p.

Finally, some eukaryotic repressors bind to DNA sequences very close to the promoter and eliminate transcription by blocking RNA polymerase access to the promoter, in a manner similar to repression in prokaryotes. Eukaryotic repressors, however, play a different role to their prokaryotic analogues because they mainly modulate the activation caused by transcriptional activators.

10.8 Regulating Gene Expression: High-Level Control Mechanisms, DNA Modification, and Epigenetics

The changes in gene activity discussed so far enable the cell to respond to fairly transient changes in conditions, or make transient changes to its state of differentiation. Development and morphogenesis can involve more permanent changes in genome activity. Implicit in our preceding discussion about the involvement of nucleosomes and chromatin remodeling in regulating gene expression is the expectation that the DNA sequence will carry appropriate target sites to make these processes gene-specific. Little is known about this aspect of the involvement of higher-order structure in transcriptional regulation, however, although it could enable coordinate regulation of multiple genes within large chromosomal domains.

DNA sequences that are potential candidates for such regulation, which must act over tens to hundreds of kilobases, include enhancers that interact with distant promoters, and may be located upstream or downstream, SAR/MAR sequences, which attach chromatin to the nuclear matrix, and LCRs, which are *cis*-acting regulatory sequences, which function, like enhancers, by binding to transcription factors with activation domains. They maintain open function across a chromosomal domain comprised of several

to many genes in a cluster that are active only during specific developmental stages. A locus control region operates sequentially with other transcription factors at *cis*-regulatory regions that are directly adjacent to each gene in clusters of related genes at different times during development. A fully assembled LCR-transcription factor complex is called an *enhancesome*. Absence of any component of the enhancesome prevents expression of the whole sequential program, and none of the genes in the cluster is activated.

The best current example of LCR-based regulation is the control of globin gene clusters in human red blood cells. The LCR in the β-globin cluster is spread over about 10 kbp between 5 and 18 kbp upstream of the first globin gene. It includes four 300 bp regulatory regions. The α-globin LCR is a region of 300 bp lying 40 kb upstream of the embryonic globin gene. The regulatory regions of the LCR contain binding sites for a number of transcription factors and LCRs probably interact with the regulators of individual globin genes to activate, enhance, and developmentally regulate their expression. It is evident that the system provides a mechanism for restricting gene expression to a specific cell lineage. The LCR, oddly enough is not required for activation of the β-globin cluster in the mouse, so there may be functional (or even experimental) differences between the two mammals.

Another example of long-term gene silencing is provided by the sequences at either side of the mating-type locus in yeast. The active chromosomal locus of the mating-type gene is called *MAT* (for mating type; see Section 2.6), but there are two additional copies of the mating-type gene, called *HML* and *HMR*, with one each side of *MAT* and located near the telomeres on each arm of chromosome III. *HML* and *HMR*, at least one of which carries a different DNA sequence to the active locus, are storage loci from which active copies are retrieved by intrachromosomal recombination during mating-type switching (Section 2.6). Under normal circumstances, the storage loci are transcriptionally silent, and are kept inactive by external silencer signals flanking the loci. The major genes involved in maintaining this silencing are four *SIR* (silent information regulator) loci, but a gene involved in maintaining telomeric heterochromatin in an inert state is also required, as is an intact histone H4.

Mutations that eliminate the activity of any *SIR* gene, that delete a *cis*-acting silencing enhancer, or mutate the N-terminal region of histone H4 all abolish silencing. Removing silencing allows simultaneous expression of both mating-type idiomorphs, so the cells behave as diploids and do not mate. In addition, in such mutants both *HML* and *HMR* become targets for mating-type switching, implying that regulation of recombination and transcription targets involve the same molecules. The SIR polypeptides form a *trans*-acting complex that acts at the *cis*-acting sites near *HML* and *HMR*, binds to other polypeptides, and interact with histones H3 and H4 to form a transcriptionally silent chromosomal domain of heterochromatin. Heterochromatin is highly condensed chromatin that is completely silent, lacking

even basal transcription as long as the silencing structure is in place. Eukaryotic chromosomes have heterochromatic regions at centromeres and telomeres, and in higher organisms whole chromosomes can be inactivated this way. The most celebrated example of this is X-chromosome inactivation in female mammals, in which one of the two X-chromosomes is inactivated and perpetuated in a heterochromatic state (the active X-chromosome is part of the euchromatin). This is an example of epigenetic inheritance, which is a heritable change in phenotype that does not result from a change in genotype.

In mammals, this can be expressed as genomic imprinting, which results from selective silencing of the expression of genes inherited from one parent or the other. Genomic imprinting is a violation of the tenets of Mendelian inheritance, which state that the parental origin of an allele has no effect on the phenotype of the progeny. Chromosomal regions silenced by heterochromatin in this way are often associated with DNA methylation. DNA methylation involves the enzymatic addition of methyl ($—CH_3$) groups to DNA, either at position C-5 of cytosine or at position N-6 of adenosine by DNA methyltransferase enzymes (= DNA methylases). In prokaryotes, DNA methylation is used to modify specific DNA sequences to protect against restriction endonucleases and in directing DNA repair systems to parental DNA strands (which are methylated) rather than newly synthesized strands (which are not methylated) to correct mistakes in replication. In eukaryotes, the major (perhaps the only) modified base is 5-methylcytosine.

The highest levels occur in plants, where up to one third of all cytosines in the genome can be methylated, and vertebrates, where the methylated sequence is CG (i.e., methylated cytosine is always followed by G on the 3′-side). The sequence -CG- is self-complementary, and in fully-methylated DNA, methylated cytosines therefore occur in pairs on opposite strands. The most efficient substrate for methyltransferases is a -CG-/-GC- pair in which only *one* cytosine is methylated, which arises each time methylated DNA is replicated. As a result, the pattern of methylated cytosines is replicated by a burst of methyltransferase activity after each cell division. Methylated-CG interacts with *trans*-acting components to alter chromatin structure and prevent transcription. DNA methylation regulates plant development by repressing transcription, and its influence is particularly evident in flower development. DNA methylation is also involved in vertebrates in coordinating gene regulation during development. Mice that are genetically deficient in the DNA methyltransferase are unable to complete development. Mutations that affect DNA methylation have very different phenotypes in fungi, plants, and mammals, which indicates that DNA methylation serves very different functions in these organisms. DNA methylation seems to be used to impose epigenetic programs on mammalian embryonic development.

The extent of DNA methylation is consistently low in fungi. Indeed, DNA methylation is not detectable in *Saccharomyces cerevisiae*. It is interesting that the classic model of animal genetics, *Drosophila melanogaster*, also lacks detectable methylated-CG in the genome. DNA methylation, however, does seem to play important roles in at least some filamentous fungi, even though genomic methylation levels are quite low compared with mammals and plants. In filamentous fungi methylation may be used primarily as a "genome defense system" for silencing repeated DNA regions. The rules governing silencing by methylation in fungi are not yet completely clear, but repeated sequences are especially susceptible. The surveillance mechanisms, on which genome defense systems depend, monitor the arrangement and content of the genome by detecting sequence homology. Duplicated genes or genes reorganized in some other way can interfere with proper genetic regulation and their expression must be prevented. For this, filamentous fungi use DNA methylation, point mutations, or both. The only repeats that escape modification are the ribosomal DNA repeats, which seem to be protected by some special feature of their location. Large duplications that result from chromosomal rearrangements are modified, but the main aim of DNA modification in filamentous fungi is to protect the genome from transposable elements.

Several types of transposable elements have been found in these fungi, but all suffer modification unless they are less than about 200 bp. The first such process discovered, in *Neurospora crassa*, was called repeat induced point mutation (RIP). A related process, named methylation induced premeiotically (MIP), was subsequently found in *Ascobolus immersus*. There is some evidence for similar processes in other filamentous fungi. RIP and MIP search the genome specifically when haploid nuclei of compatible mating types are in a common cytoplasm. Neither of these processes operate after this dikaryotic stage (i.e., MIP and RIP detect sequence duplications that are similar enough, and long enough, to allow them to pair in the haploid genomes of nuclei *before they undergo karyogamy* and meiosis). Mutagenesis by GC to AT transitions is the major consequence of RIP, but remaining cytosine bases in the affected sequences are also frequently methylated after RIP. The mutation process may involve cytosine methylation followed by deamination. Thus, RIP causes both genetic (the mutational) and epigenetic (the methylation) changes. The mutations caused by RIP usually inactivate affected sequences completely, although the extent of mutation is variable.

The analogous system in *Ascobolus*, MIP, causes methylation without mutation. A gene attacked by MIP in *Ascobolus* produces no mRNA because methylation by MIP interferes with transcription elongation. Silencing by MIP can be reversed if methylation is prevented in growing hyphae of *Ascobolus*. Methylation, however, is much more variable in filamentous fungi than it is in mammals. Methylation in fungi is not confined

to symmetrical -CG-/-GC- sites, and a methylated sequence can coexist with an identical unmethylated version. It may be that some distinctive feature of the chromatin is more significant than the absolute degree of methylation. Repeat-induced methylation also occurs in the basidiomycete *Coprinus cinereus*, but the sequences are rarely silenced because the methylation in this case is sparse.

Another form of silencing in vegetative hyphae of *Neurospora* is quelling. Quelling is specifically induced when transforming DNA (also known as transgenes or ectopic DNA sequences) that is homologous to endogenous DNA sequences is introduced into *Neurospora*. The transforming DNA inhibits expression of the homologous gene, even when the two sequences are not linked. Very short regions of homology are required to induce quelling; as little as 200 bp will suffice, as long as it is from the coding region of a gene. Transgenes containing the promoter region only are not effective. Quelling results in reduced levels of mature mRNA, but the level of primary transcript is not significantly reduced, so quelling must act post-transcriptionally. It does not involve methylation or direct DNA–DNA interactions, but it does involve a *trans*-acting molecule that is expressed through the cytoplasm. It is probable that the silencing agent is a sense RNA that participates in RNA–RNA or RNA–DNA pairing.

Quelling appears to be related to a range of homology-dependent gene silencing (HDGS) processes that occur naturally in eukaryotes. HDGS first became evident when experiments on plant transformation resulted in some transgenes inducing self-silencing as well as silencing homologous transgenes and endogenous sequences. Plants exhibit two forms of HDGS: transcriptional gene silencing (TGS) is caused by suppression of transcription, and posttranscriptional gene silencing (PTGS) is due to mRNA degradation. So far, quelling is the only posttranscriptional gene-silencing process that has been found in fungi.

These plant and fungal phenomena are related to transvection, which was discovered in the fruit fly, *Drosophila melanogaster*. Homologous chromosomes are paired in somatic cells of *Drosophila*, the clearest, and most classic, expression of this being the giant polytene chromosomes of the salivary glands. This pairing of homologues influences gene expression *in vivo*, and disruption of it can influence development. A gene that exhibits transvection has its function altered by homologue pairing; transvection can lead either to gene silencing or activation. The mechanism can involve direct DNA–DNA contact or pairing through intermediary factors. The test for transvection is to show that disrupting somatic (or meiotic) pairing between, for example, two alleles of a gene is sufficient to alter the phenotype. The test has proved positive with several genes, and the developing picture is that homologue pairing is a modulator of genome function. There is no decisive indication of a mechanism, however, but rather a collection of likely candidates. These include: (1) pairing allowing enhancers to act on the homologous sequence, (2) propagation of chromatin structure from one

homologue to the other, (3) pairing of special sequences leading to the assembly of a silencing chromatin structure, (4) pairing allowing the concentration of a special RNA to trigger silencing, (5) pairing generating a chromosomal topology that augments gene expression.

Evidence of transvection has been found in the brief diploid phase of *Neurospora*. Strains with only one copy of an ascospore maturation gene (*asm*-1), or those strains with two copies located at nonallelic sites, produced only a token number of mature ascospores, implying that homologous pairing of *asm*-1 alleles is required for full expression of this gene. It is interesting that paired alleles supported maturation of spores bearing a wild-type allele even if the other allele had a frameshift mutation rendering it nonfunctional in the vegetative phase. Results with several other *Neurospora* genes suggest that the fungus might use transvection generally to control expression of development-specific genes. These homology effects force one to recognize that unusual forms of gene regulation involving DNA–DNA, RNA–DNA, and RNA–RNA interactions at the chromosomal level may well prove to be important regulatory processes. Bearing in mind how unsuitable ascomycetes are for studies of dikaryosis, it would be interesting to know whether transvection is important in basidiomycetes, in which stable haploids, heterokaryons, dikaryons, and diploids can be compared at all stages in the life cycle.

10.9 Posttranscriptional Regulation: Spliceosomes, Proteasomes, and Protein Networks

Although most regulatory mechanisms control transcription, there are several posttranscriptional events that offer the opportunity for regulation. These include RNA splicing, RNA stability, mRNA editing, trafficking between nucleus and cytoplasm, protein synthesis, and protein stability. As indicated earlier (Section 10.5), most of the pre-mRNA processing reactions occur on the nascent transcripts associated with the transcription machines. Particular sequences in the primary transcript define the borders between introns and exons and are recognized by a range of *trans*-acting factors that make up the spliceosomes.

The spliceosome comprise small nuclear ribonucleoproteins (snRNPs), composed of uridine-rich snRNAs together with a particular collection of RNA-binding proteins, and various nuclear proteins. Many of the latter have a domain rich in serine-arginine dipeptides as well as one or more RNA-binding domains. They are called *SR proteins* because S and R serve as the single-letter abbreviations of serine and arginine, respectively. The RNA-binding specificity of SR proteins can provide them with the ability to determine alternative splicing patterns. While they are in the nucleus, transcripts synthesized by polymerase II are associated with a group of

abundant RNA-binding proteins called heterogeneous nuclear ribonucleoproteins (hnRNPs). These are involved in virtually every aspect of pre-mRNA processing, transport, and translation. The tissue-specific expression of SR-proteins and hnRNPs is likely to be critical to the fate and function of the transcription product, but it is a largely unexplored aspect of gene regulation. The hnRNP and SR-proteins that shepherd the transcript from first synthesis by the transcription machines also contribute signals for export to the RNPs that translocate mRNA through the nuclear pore complexes (NPCs) that perforate the nuclear membrane.

The population of RNAs in a cell changes over time; messengers are synthesized and messengers are degraded. Most eukaryotic mRNAs are modified by the addition of a poly-A tail at their 3′-ends. This is not encoded in the DNA template of the transcript and it can vary in length from less than 20 to more than 200 adenines. Cytoplasmic enzymes gradually remove poly-A tails, and once the tail has been removed the rest of the mRNA is degraded. Other things being equal, mRNAs with many adenines in the tail will have a longer life in the cytoplasm than those with few adenines. Gene functions that have to be rapidly changed require short-lived mRNAs; otherwise, the mRNA will remain in the cytoplasm long after transcription of the gene has been repressed. Thus, the working lifetime of the messenger is an important aspect of the regulatory strategy of any gene. It is usually measured as a "half-life," which is the length of time necessary to reduce the population of molecules to half its original value. Yeast mRNAs have half-lives averaging 10–20 minutes, although some mRNAs have half-lives of only 1 minute and others about 35 minutes; in mammals, the average can be several hours.

Messenger RNAs can include untranslated sequences that influence their translation. Control of translation is most commonly applied at the initiation phase because this is the rate-limiting step. The 3′ poly-A tail is also involved in the initiation of translation, where a poly-A-binding protein is one of the important initiation factors needed for binding of the small ribosomal subunit near the 5′-end of the messenger. The messenger circularizes to enable this recruitment of the ribosomal subunit. Because it requires a synergistic interaction between 3′ poly-A tail and 5′-cap structure, circularization probably ensures that only properly processed mRNA molecules are translated. Regulation of initiation factor function, by phosphorylation or cleavage, can impose global control on translation and thus total protein synthesis. In addition, translation of specific mRNAs can be regulated by *cis*-acting elements on the mRNA: the untranslated regions or UTR sites. Because circularization of the mRNA is so important for initiating translation, UTRs located at the 3′-end of the mRNA are commonly used as *cis*-acting elements controlling mRNA translation and/or localization.

Transcript localization has mostly been studied in animal cells. We have mentioned (Section 2.6) the best-known fungal example, which is the local-

ization of *Ash1* mRNA to the bud to repress mating type switching, but it would be surprising if it were not used extensively in filamentous fungi because the fungal hypha is an ideal candidate for such compartmentalization. Two characterized (animal) examples reveal alternative strategies. Linkage to molecular motors for directional transport on cytoskeletal tracks can localize transcripts. In contrast, transcript localization can be achieved by generalized transcript degradation combined with localized protection. Transcript localization and translational regulation may be intimately connected because for certain messengers only the localized mRNAs are translated; the unlocalized transcripts are translationally repressed.

Once the polypeptide has been synthesized, various posttranslational modifications subsequently affect protein function. These include binding of substrates, coenzymes, and metabolic products, (called ligands), as well as covalent modification reactions (e.g., oxidation, acetylation, or phosphorylation). Such posttranslational modification is usually much more rapid than transcriptional control and therefore suits situations that require a rapid response to a stimulus (e.g., in signaling cascades) (Section 10.11).

The lifetime of the polypeptide product is another consideration, and there are many enzyme systems that destroy proteins. For example, yeast cells contain more than 40 peptidases, but most are involved in specific protein processing rather than general protein degradation because only seven have been found in lysosomal vacuoles. Many polypeptides that have short half-lives contain one or more regions rich in the amino acids proline (single character code letter = P), glutamic acid (E), serine (S), and threonine (T); obviously, they are called PEST regions. Proteins with long half-lives do not have PEST regions, so they presumably identify the polypeptide as a target for degradation. Such motifs can identify proteins for uptake for degradation by lysosomal proteinases or for binding by peptide recognition protein for disposal by a nonlysosomal mechanism. In general, nonlysosomal mechanisms are used to degrade proteins with relatively short half-lives, whereas the longer-lived proteins are degraded in lysosomes. One of the nonlysosomal mechanisms attaches chains of ubiquitin as a marker to target other proteins for degradation; the process is called *ubiquitination*.

Ubiquitin is a highly conserved polypeptide, containing 76 amino acids, that serves as a tag for the recognition of proteins for proteolysis by the multicatalytic proteasome in yeast, as well as higher eukaryotic cells. Each proteasome complex contains many subunits and multiple catalytic centers. Cytoplasmic proteins that are old or damaged, or are candidates for regulated destruction (e.g., cell-cycle proteins or transcription factors), are modified by the addition of a chain of several to many ubiquitin molecules attached as a linear or branched polyubiquitin chain, which is recognized by the proteasome. The targeted protein is then threaded into the inside of

the proteasome and reduced to peptides by the internal proteolytic enzymes.

Plasma membrane proteins (receptors and transporters) in yeast, and in animal cells, are also ubiquitinated at the cell surface in response to ligand binding as a signal for internalization and downregulation. In these cases in yeast, though, ubiquitination triggers degradation in the lysosome rather than in the proteasome. An example is the ubiquitin-dependent internalization of mating-type pheromone receptors in *Saccharomyces cerevisia*e. When the receptor is activated by pheromone binding, it is phosphorylated and subsequently ubiquitinated. A single ubiquitin molecule is sufficient to promote rapid internalization, followed by lysosomal degradation. This type of ubiquitin modification is different from the polyubiquitination that is required for recognition of proteins targeted for degradation by the proteasome, so there are at least two ways in which interaction with ubiquitin is used to modify other proteins.

Interactions between proteins are an essential element in each one of the regulatory phenomena we have described so far, from transcription to protein degradation, and it would probably not be an overestimation to claim that every cellular process depends on polypeptide–polypeptide interaction. This dependence ranges from the need to create structures (e.g., cytoskeleton, nuclear scaffold, division spindle, nuclear pores, centrosomes, kinetochores, etc.) through to transient protein–protein interactions that control and regulate so many cellular reactions. As more and more genome-sequencing projects are completed interest in how the genome is reflected in the phenotype has shifted from the particular toward the holistic in an attempt to approach understanding of the interplay of gene products with other molecules in a cell.

One aspect of this is the large-scale identification and display of protein interactions that give rise to protein interaction maps that represent the network of interactions between proteins (Fig. 10.4). The technique that has enabled large-scale analysis of protein interactions more than any other is the yeast two-hybrid system. This method allows proteins to be assayed for interaction simply by examining the growth of yeast colonies on a plate. The method uses the fact that many eukaryotic transcription activators have two functional domains, one that directs binding to a promoter DNA sequence and one that activates transcription.

The two-hybrid technique exploits the facts that the DNA-binding domain of an activator is incapable of activating transcription unless associated, physically though not necessarily covalently, with an activating domain. The activation domain of one activator can be associated with the DNA-binding domain of a second to create a functional transcription activator in yeast. In a two-hybrid experiment in practice, the protein of interest is fused to a DNA-binding domain and inserted (transfected) into a yeast cell that has a reporter gene under the control of the *cis*-acting element of this same DNA-binding domain. This hybrid protein cannot acti-

vate transcription on its own, but it can be used as "bait" to screen a clone library that contains cDNA clones fused to an activation domain. Any cDNA clones in the library that encode proteins able to interact with the bait will consequently assemble a complete transcription activator as a result of that interaction, and the reporter gene will be expressed.

Some proteins are not suited to this approach, and there are several ways in which false-positives can arise; however, these problems aside, the two-hybrid approach has generated a wealth of information about potential protein–protein interactions. When compiled into maps that depict the network of interactions between proteins, the maps provide a rough outline of the complexity of protein associations, but they also depict potential signaling pathways, interactive complexes, and clues to the function of previously uncharacterized proteins. Proteins of similar function can be classified together to generate a map that shows the interactions between proteins assigned to functional classes in yeast. Even these "simplified" maps readily indicate the complexity of the 8000–12,000 protein interactions that are thought to occur in the yeast cell.

Protein interaction studies have functions other than simply to remind us of their complexity. The main goal of interaction studies is to learn about individual proteins. What are their likely functions, potential partners, or to which complexes do they contribute? Interaction networks can contribute to this because it has been established that more than 70% of all interactions between experimentally characterized proteins in the yeast network occur between two partners of the same functional class. If interactions were randomized, only 12% of all them would belong to the same class. As a result, the functional category of an unknown protein (i.e., a protein of unknown function, identified, perhaps, as an ORF in a genome sequence) can often be assigned by identifying its partners in an interaction network.

As one of the groups working in this area has described it: "If protein *X* (uncharacterized) is found to interact with protein *Y* and protein *Z*, and both *Y* and *Z* are components of the RNA-processing machinery, then it is quite likely that protein *X* is also involved in RNA processing, perhaps as part of a complex with *Y* and *Z*." Nobody has yet produced a comprehensive protein interaction map of any cell type, and we are far from being able to display the spatial and temporal expression patterns that must be considered if we are to understand how cellular differentiation and tissue morphogenesis are regulated. Truly holistic studies, which incorporate genetical, biochemical, physiological, morphological, and temporal information, need far more experimental analyses and will generate orders of magnitude more data than we have at the moment. This, in turn, will require improved data-management software and new tools for visualizing complex information; a new approach to integrated science. For the moment, we have to turn away from holistic developmental biology and virtual visualization and return to the real world, where the fungi grow!

10.10 Shape, Form, and Differential Gene Expression

A general feature of development in eukaryotes is that only a small proportion of the genome is associated with any particular morphogenetic process. The emphasis in morphogenetic gene regulation is on differential expression of activity rather than on large-scale replacement of one set of gene products by another. This is also true for the fungi despite their having a generally smaller genome size than other eukaryotes.

Fungal examples of differential gene expression have been revealed in relation to the comparison between homokaryotic and dikaryotic phenotypes in *Schizophyllum*, the transition from vegetative state to fruit body formation in *Coprinus*, perithecium formation in *Neurospora* and *Sordaria*, sclerotium development in *Sclerotinia*, and sporulation in *Saccharomyces*. In the yeast example, only 21–75 genes out of the total genome were found, by classical genetic analysis, to be specific to meiosis and ascospore formation. Comparisons between fruiting and nonfruiting cultures like this involve cultures of similar age, which, for environmental or genetic reasons, differ in their ability to undergo a morphogenetic change. This approach is as exclusive as is technically possible and seeks to identify genes on which the morphogenetic change is causally dependent. It shows that there are relatively few of these.

A contrasting all-inclusive approach identifies the subset of the genotype that contributes to the morphogenetic change. For example, when the mRNA sequences that accumulated in conidial cultures of *Aspergillus nidulans* that had germinated for only 16 hours (assumed to represent purely vegetative hyphae) were compared with conidiating cultures grown for 40 hours, it was found that 11–18% of sequences occurring in sporulating cultures were not detectable in vegetative hyphae, and that 6% of the unique sequences were expressed during conidiation. This type of comparison shows the sum total of differences between cultures grown for 16 and 40 hours together with differences due to the differentiation associated with conidiation together, possibly, with other age-related differences (e.g., secondary metabolism), which may have no relation to conidiation. The gene subsets involved in different morphogenetic events evidently differ by a fairly large minority of the expressed genes (this is what is meant by "differential gene expression"), but do have shared components. Even very different pathways of morphogenesis may share aspects of cell differentiation (e.g., particular parts of primary metabolism, cell inflation, wall thickening, accumulation of metabolites, etc.). The problem is to distinguish between the causal and the merely contributory features.

In terms of microbial growth kinetics the term *balanced growth* describes the growth that occurs when all nutrients are available in sufficient quantity and the microbial cells can synthesize all of their components in balance. Unbalanced growth occurs when some limitation, whether nutri-

tional or environmental, adversely affects synthesis of one or more of the cellular components, yet growth persists and the cells that are formed are abnormal in the sense that they differ from those produced by balanced growth. The relevance to the present discussion is that the growth pattern of a differentiated cell is "unbalanced" in comparison with the growth pattern of an undifferentiated vegetative cell. The direction, progress, and extent of the imbalance are precisely what define the state of differentiation, yet there are numerous ways in which unbalanced growth can be precipitated.

By analogy, therefore, perhaps there are numerous ways in which a state of differentiation can be initiated. The master genetic control elements may be involved in defining and providing for the events that *could* take place, rather than being causally involved in what *will* take place. Causality may rest with altered temperature, pH, nutrition, and so on, which expedite change. The manner of the change may depend on the past history of the cell and the future avenues for change, which that history has made possible. In a crude analogy, bricks may be necessary to build a house, but the manufacture of bricks does not determine the shape of the house, nor even that it should be built. In the absence of bricks, of course, a house could be built of timber.

10.11 Yeast–Mycelial Dimorphism

Several fungi have the ability to switch their growth pattern between a cellular yeast form and a filamentous hyphal form in response to environmental cues. This is called *dimorphism*, and it is observed in plant pathogens like *Ustilago maydis* and human pathogens like *Candida albicans*. Nonpathogens also show dimorphism, including species of *Mucor* and even the "classic yeast," *Saccharomyces cerevisiae*. Two different sorts of filamentous growth occur. In the pseudohyphal mode, yeast cells become elongated and fail to separate after cell division, remaining attached to form chains of elongated cells. On the other hand, a true hyphal filament may be produced.

The ability to switch between yeastlike and filamentous forms has been correlated with virulence in pathogens because dimorphic transitions are often required at some stage in the infection process. For example, filaments can be invasive, penetrating solid tissues and beneath the surface to which the infecting cell form is initially attached. A yeast form is equally easily distributed in fluid flows and can enable the pathogen to widen its invasion of the host by transport through the circulatory systems in animals and plants alike. The genes that control dimorphism in pathogens have been the focus of many studies because they could offer new, and specific, targets for antifungal agents. The dimorphic switch is triggered by various signals *in vitro*, and many of the responses can be related to the normal interactions

between the fungus and its environment *in vivo*. In large measure, therefore, the study of dimorphism has become a study of the signaling pathways that connect the external environment with a change in cell differentiation. It has become evident that these pathways are broadly conserved within the fungi.

Dimorphism has been studied in two particularly well-known pathogenic fungi (i.e., *Candida albicans* and *Ustilago maydis*), which infect mammals and maize, respectively. Several signal transduction pathways have been defined in both fungi, and they have proved to be similar to pathways involved in the pseudohyphal differentiation that results when some strains of *Saccharomyces cerevisiae* are starved of nitrogen. Although different fungi use strikingly similar signaling pathways to respond to environmental cues, however, the outcome of the signaling events can be very different. All organisms respond to cues from the environment outside the cell. These signals are transduced from the cell surface to the interior of the cell, and ultimately to the nucleus, resulting in altered gene expression. The consequential change in the pattern of protein activities then results in the cellular response to the external environment.

The literature on this topic is vast and not entirely genetically relevant, so some sweeping generalizations are called for: (1) any given cell contains multiple signaling pathways, each of which responds to a distinct signal that is transduced to give a specific response; (2) a given signaling component can be used in more than one pathway to respond to different signals; (3) different organisms may use the same pathways to respond to the same signal, but some of the components may be used differently; (4) the environmental signal is perceived at the cell surface by various types of receptor. In eukaryotes, the central component of these signal transduction pathways is the mitogen-activated protein kinase (MAPK) cascade. Three highly conserved protein kinases make up this cascade: MAPK (also known as extracellular signal-regulated kinase, ERK), MAPK kinase (MAPKK, also known as mitogen-activated, ERK-activating kinase, MEK), and MAPK kinase-kinase (MAPKKK, also known as MEK kinase, MEKK). Sequential activation of these kinases by phosphorylation is the most vital part of the transduction and *amplification*, of the signal through the cascade. MAPK is activated by MAPKK, which is in turn activated by MAPKKK. The latter is activated by the signal receptor. Following activation of the MAPK cascade, activated MAPK generates an output signal (e.g., a transcription activator).

Receptors used in different pathways may be G-protein-coupled (serpentine or seven-trans-membrane) receptors, His-Asp phospho-relay sensors, or integral membrane proteins. G-proteins have essential roles in sexual and pathogenic development. For example, they are part of the mating-type pheromone-signaling cascade in both ascomycetes and basidiomycetes. In addition, G-proteins affect a number of developmental and morphogenetic processes.

In both haploid and diploid strains of *Saccharomyces cerevisiae*, starvation for nitrogen, and possibly other stresses, activate a MAP kinase cascade, which has the transcription factor Ste12 as its final target. Pseudohyphal growth is the eventual outcome. The genes concerned are also essential for mating pheromone response, emphasizing the involvement of mating-type signaling pathways in the dimorphic transition. A cAMP-dependent pathway operates in parallel, and specifically includes a protein kinase called Tpk2. The two pathways together modulate expression of a gene that encodes a cell-surface protein required for pseudohyphal and invasive growth. The MAP kinase cascade also controls complexes that affect both the budding pattern and cell elongation of yeast cells. Another gene product that stimulates filamentous growth in *S. cerevisiae* is Phd1; although its mode of action is unknown, it may be regulated by a cAMP-dependent protein kinase. A homologous protein, Efg1, is found in *Candida albicans*, and is regulated in this way. Efg1 is required for the formation of true hyphae.

Dimorphism of *Ustilago maydis* is governed by the *a* and *b* mating-type loci (Section 2.8). The *a* factor is necessary for conjugation tube formation and the *b* locus produces true hyphal filamentous growth. Filamentous growth is dependent on stimulation by the mating pheromone pathway and on a panel of genes whose expression is strictly limited to the filamentous phase and is directly or indirectly regulated by the mating-type factors. This model of mating-type regulation has encouraged a search for molecular switches of similar sorts in organisms in which dimorphism is not part of the sexual cycle. The search is not straightforward, however, because of the wide range of metabolic and environmental factors that influence or govern dimorphism. Even in *Ustilago maydis* the dimorphic switch responds to environmental and metabolic conditions. An acidic medium is sufficient to induce development of the mycelial form, which suggests that growth at low pH overcomes the control processes governed by the *b* mating-type factor. The cAMP-dependent protein kinase-signaling pathway is involved metabolically in controlling morphogenesis in *U. maydis*. Disruption of the gene-encoding adenylate cyclase results in a constitutively filamentous phenotype, where budding is restored by growth in the presence of cAMP.

Because such relatively unspecific environmental/metabolic effects can be seen in an organism in which the dimorphism is known to be regulated by identifiable master control genes (e.g., the mating-type factors), it is not surprising that there is vigorous debate about causality in those organisms that lack an obvious master control genetic element. Unspecific effects are readily demonstrable: dimorphic alterations during temperature shifts in *Paracoccidioides brasiliensis* are preceded by significant changes in protein synthesis during the yeast to mycelium transition (temperature downshift 36–26°C), although there are few changes during the mycelium to yeast differentiation (26–36°C). More specific differences can also be

found (e.g., promotion of differentiation of the dimorphic yeast *Yarrowia lipolytica* from yeast to mycelium by a gene product) that stimulates Golgi secretory function, but these are all associations rather than causes. Even in more extensively studied fungi, there is considerable debate over the relative importance of differential gene expression and the pattern of metabolism.

Another example is dimorphism in the zygomycete *Mucor.* The chemistry of the wall is similar in both phases, but what distinguishes them is the way in which the wall is synthesized: isodiametric in the yeast form, apical and vectorial in the hyphae. Various enzyme activities and physiological processes alter during the dimorphic change, but none seem to be strictly causal. Cyclic AMP and other signaling molecules, as well as enzymes governing their intracellular concentrations, also show consistent dimorphism-related patterns of change. Cytoskeletal components and their protein kinase regulators are similarly involved in apical growth, but a causal link is again lacking. A similar story can be told for dimorphism of *Candida albicans*, in which gene expression has been extensively studied; however, the studies reveal complex alterations in gene expression during the dimorphic transition with most genes examined showing transient or persistent increases or decreases in mRNA levels. Further complication is added by strain- or medium-dependence of morphogenesis-specific gene expression of two chitin synthase genes (CHS2 and CHS3) and three aspartyl proteinase genes that, in the affected strain and/or effective medium, are transcribed preferentially in the hyphal form.

In general, yeastlike cell morphology results from a particularly patterned interaction of polarized growth and cytokinesis, combined with subsequent division of the cell wall to separate daughter cells. Imbalances of the equilibrium of these two central processes lead to formation of morphological variants. There are numerous physiological examples that could be cited. For example, when osmotic stress is applied to filamentous water molds their growth can become disorganized with weak and malformed hyphal walls, and the filaments will take on shapes like budding yeasts. Some chemostat-grown fission yeasts cultured at low dilution rate (i.e., under nutrient limitation) divide asymmetrically, yielding daughter cells of unequal volume. Fission yeast shaped like round-bottom-flasks were induced by treatment with aculeacin A, which inhibits β-glucan synthesis, although α-glucan synthesis continues apparently normally. This suggests that imbalanced glucan synthesis gives rise to the aberrant cell shapes, and everything else is a consequence of the cell coping with the deformed cell shape caused by that unbalanced wall assembly. This idea of an altered equilibrium, or imbalance, causing morphological change(s) could be a valuable generalization as a mechanism that could drive a wide range of fungal cell differentiation processes.

10.12 Conidiation: Translational Triggering and Feedback Fixation

Conidia are asexual spores of ascomycetes that can survive in a dormant state for longer periods than vegetative hyphae. They arise on specialized hyphae called conidiophores. The spores themselves are essentially rounded off hyphal segments that are often more or less spherical in shape, which detach for dispersal. The formation of conidia by surface cultures of *Aspergillus nidulans* occurs after about 16 hours hyphal growth, which is a period of vegetative growth required to make the cells competent to respond to the induction process. Induction requires exposure to air and is probably a reaction to cell-surface changes at air–water interfaces. After induction, some mycelial hyphae produce aerial branches that become conidiophore stalks.

The cell from which the branch emerges is the conidiophore foot cell (Fig. 10.5), which is distinguishable from other vegetative cells by having a brown pigmented secondary wall thickening on the inside of its original wall. The stalk grows apically until it reaches a length of about 100 µm when the apex swells to form the conidiophore vesicle, which has a diameter of

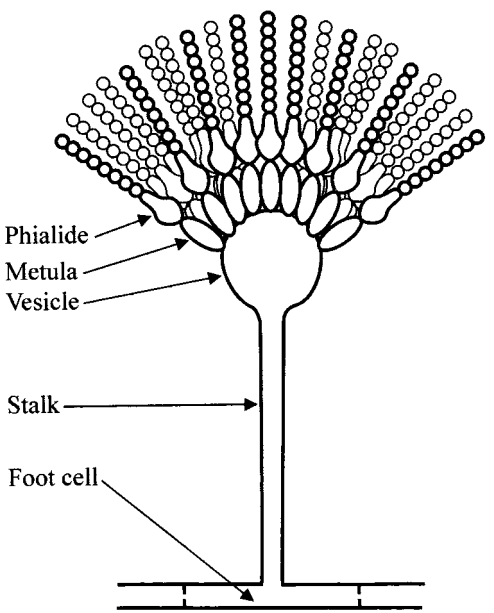


Fig. 10.5. Diagrammatic structure of the conidiophore of *Aspergillus nidulans*. (Figs. 10.5–10.8 based on Figs. 5.11–5.16 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

about 10 μm. A single tier of numerous primary sterigmata, called *metulae*, then bud from the vesicle, and secondary sterigmata—the phialides—bud from the exposed apices of the metulae. The phialides are the stem cells, which then undergo repeated asymmetric divisions to form the long chains of conidia that are approximately 3 μm in diameter (Fig. 10.5).

Classical genetic analysis, by isolation and analysis of mutants, has been used to establish the basic genetic outline of this process. By comparing mutation frequencies at loci affecting conidiation with those for other functions it was estimated that between 300 and 1000 loci are concerned with conidiation. Analysis of mRNA species indicates that approximately 6000 are expressed in vegetative mycelium, and an additional 1200 are found in cultures that include conidiophores and conidia, 200 of these additional mRNAs being found in the conidia themselves. As mentioned earlier, though, this method does not distinguish conidiation-specific mRNAs from those coincidentally associated with conidiation.

Only about 2% of mutants of *A. nidulans* that lacked conidia (aconidial mutants) have defects in stages concerned with conidiophore growth and development. By far the majority (83%) are defective in the preconidiophore stage, and 15% are affected in conidium germination or pigmentation. Eighty-five percent of conidiation mutants are also defective for vegetative hyphal growth. Assuming all mutations have an equal chance of isolation, these proportions suggest that attaining competence involves the largest number of gene functions. Of the few genes that seem to determine conspicuous developmental events in conidiophore morphogenesis, two in particular play key roles. These are the "bristle" (*brlA*) gene, which has defects in vesicle and metula formation, and "abacus" (*abaA*), in which conidia are replaced by beaded lengths of hypha. These two loci are each represented by 30 or more mutant alleles, and no other mutations have been identified that affect these stages of conidiophore morphogenesis.

Studies of temperature-sensitive mutants, combined with epistatic interaction and RNA transcript detection studies, have indicated that *brlA* is required during vesicle, metula, and phialide stages, and *abaA* during conidial budding from the phialide. A third gene with regulatory properties is *wetA*, which is defective at an early stage of spore maturation. Conidia of *wetA* lack pigment and hydrophobicity; they fail to express a range of spore-specific mRNAs, and autolyse after a few hours. The *wetA* gene transcript is lacking in *brlA* and *abaA* mutants. Gene expression patterns and epistasis between the genes in double mutants suggests that these three genes function in the order *brlA* → *abaA* → *wetA*. There are many other *A. nidulans* mutations that affect a variety of specific functions in sporulation, but these three genes, *brlA*, *abaA*, and *wetA*, seem to be the key control elements.

A striking feature of the mutational analysis of conidiophore development in *A. nidulans* is that each phenotype is represented by mutation in just one locus. This may suggest that the genes that are isolated in mutation

analysis are regulators, which integrate the expression of other genes that are not themselves specific to conidiation. Molecular analyses support this interpretation. The amino acid sequence of the *brlA* product contains zinc fingers near the carboxy-terminus, indicating that *brlA* encodes a transcription factor, which is required for activation of transcription of developmentally regulated target genes. The regulatory network, which has emerged from the studies we describe, is illustrated in Fig. 10.6.

Phenotypes of some *brlA* mutants that have only partially lost function, and in which target genes show varied effects out of proportion to the loss of *brlA* function, suggest the *brlA* product has different affinities for different target genes. The *brlA* locus consists of overlapping transcription units: the downstream unit is designated *brlA*α and the upstream unit *brlA*β. The two share the same reading frame for most of their length, but *brlA*β has an additional 23 amino acid residues at the amino-terminal end of that reading frame, and its transcript also possesses an ATG-initiated reading frame of 41 amino acid residues (called μORF) near its 5′ terminus. The two transcription units are needed for normal conidiophore development but the two BrlA peptides they encode can substitute for each other. Their functional difference seems to be in the very earliest stages of the initiation of development. The *brlA*β transcript can be detected in vegetative hyphae, but the BrlA peptide is not translated from the transcript because translation initiation at μORF represses translation from the downstream (BrlA) reading frame. Thus, the competent hypha is primed to undertake conidiophore development, with only this translational repression maintaining vegetative growth and preventing irreversible activation of the conidiation pathway.

Activation of the conidiation pathway in this way has been called *translational triggering* because if the repression caused by μORF can be overcome, the *brlA*β transcript will be translated and BrlA will activate conidiation. The translational trigger may be a way of making development sensitive to the nutritional status of the hypha, as nitrogen limitation (a common environmental signal for initiation of sporulation) reduces aminoacyl-tRNA pools, which could disturb translational regulation by μORF. Activation of *brlA* depends on a gene called *flbA* that encodes an mRNA that is expressed throughout the *A. nidulans* asexual life cycle. The sequence encodes a polypeptide with some similarity to a *Saccharomyces cerevisiae* protein that is required by yeast cells to resume growth following prolonged exposure to yeast-mating pheromone *a*. The *flbA* protein is thought to contribute to the signaling pathway in *Aspergillus* that distinguishes between continued vegetative growth and conidiophore development.

Activation of *brlA* is therefore seen as the first step in conidiophore development, and its product in turn activates a panel of conidiation-specific genes among which are *rodA* (encodes a hydrophobic component of the conidium wall), *yA* (encodes a *p*-diphenol oxidase (= laccase) respon-

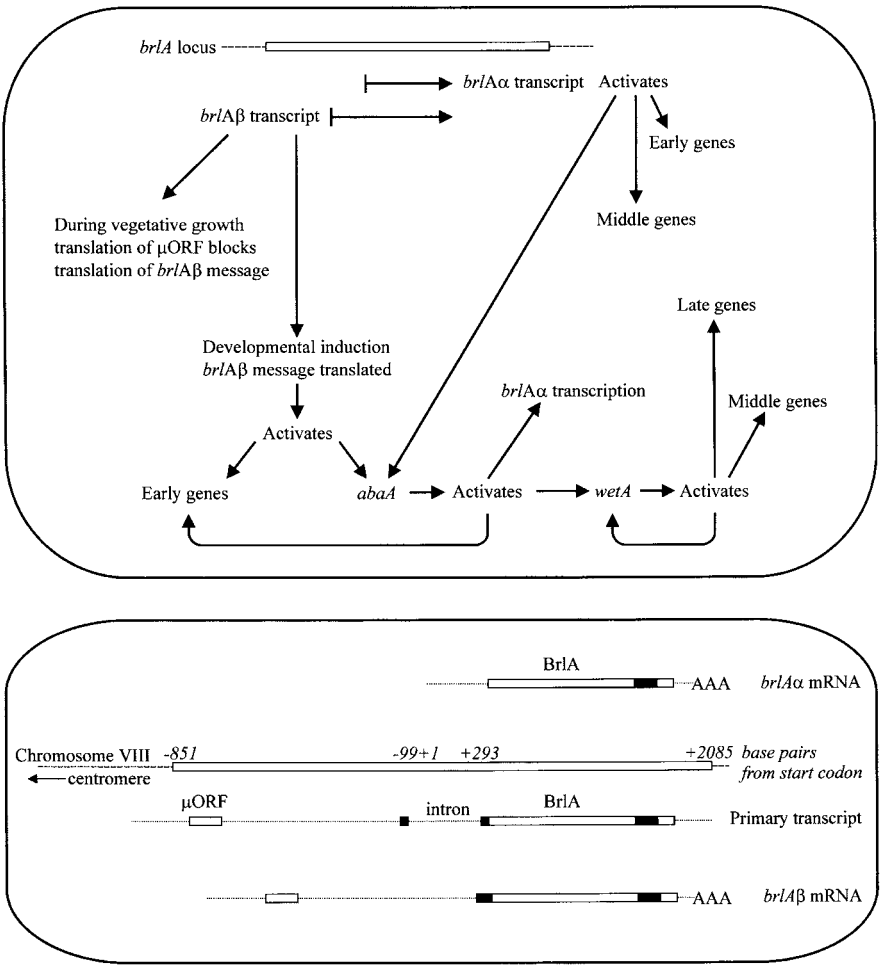


Fig. 10.6. Summary of the genetic regulatory circuit for conidiophore development in *Aspergillus nidulans* (upper panel). The lower panel shows the structure of the *brlA* locus of *Aspergillus nidulans*. The *brlA*α mRNA is shown at the top, alongside a clear box that represents the BrlA segment of chromosome VIII. The *brlA*β primary transcript and mRNA are shown in the lower part of the figure. The *brlA*α sequence is a single exon encoding a Cys_2-His_2 zinc finger polypeptide (location of the zinc fingers shown as a shaded box within BrlA). The *brlA*β sequence contains one intron. The polypeptide encoded by *brlA*β contains an additional 23 amino-terminal residues (corresponding sequence shown as a black box) and the transcript has a short upstream Open Reading Frame (μORF), which regulates translation of *brlA*β. (Figs. 10.5–10.8 based on Figs. 5.11–5.16 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

sible for conversion of yellow spore pigment to green), and, directly or jointly with *medA*, the next regulator, *abaA*. The *abaA* product is also a transcription factor that enhances expression of *brlA*-induced structural genes. The *brlA* and *abaA* genes are reciprocal activators because *abaA* also activates *brlA*. Of course *brlA* expression must occur before *abaA* can be expressed, but the consequential *abaA*-activation of *brlA* reinforces the latter's expression and effectively makes progress of the pathway independent of outside events. The *abaA* product also activates additional structural genes and the final regulatory gene, *wetA*, which activates spore-specific structural genes. Because *brlA* and *abaA* are not expressed in differentiating conidia, *wetA* is probably involved in inactivating their expression in the spores. Expression of *wetA* is initially activated in the phialide by sequential action of *brlA* and *abaA*. There is, however, evidence that *wetA* is autoregulatory. Positive autoregulation of *wetA* maintains its expression after the conidium has been separated (physically or cytologically) from the phialide.

Spatial organization of gene expression of this sort is also imposed upon the core regulators by the genes *stuA* and *medA*. Mutants in *stuA* form diminutive (*stu*nted) conidiophores with unthickened walls. This locus is classified as an auxiliary regulator because a number of conidiophore-expressed transcripts are missing in *stuA* mutants. Medusa (*medA*) mutants form conidia on top of multiple layers of metulae; these mutants are also sterile and unable to form cleistothecia. The *stuA* gene is complex: two transcripts are produced from distinct transcription start signals, both having short open reading frames (mini-ORFs) in their leader sequences. There is also some evidence for translational regulation of *stuA* expression. Both *stuA* transcripts increase in concentration by a factor of about 50 when cells become developmentally competent, and there is an additional 15-fold increase in *stuA* expression (which requires *brlA* activity) following developmental induction. The *medA* gene interacts with *brlA*, but it is not yet clear how; however, the *medA* transcript level declines following developmental induction.

Nonregulatory development-specific genes have been categorized into four classes on the basis of transcript accumulation in strains carrying mutations of the regulators. Class A genes are involved in early development and are activated by *brlA* or *abaA* or both, but independently of *wetA*. Class B genes are involved in late (spore-related) functions and are activated by *wetA* independently of *brlA* or *abaA*. Genes put into classes C and D are thought to encode phialide-specific functions, and their activation requires the combined activity of all three regulators.

The genetic structure revealed in this analysis is significant because it demonstrates that the conidiophore developmental process is naturally divided into sequential steps. Translational triggering exposes a mechanism that can relate a developmental pathway to the development of competence on the one hand, and to initiation in response to environmental cue(s) on

the other. Further, the reciprocal activation, feedback activation, and autoregulation seen in the core regulatory sequence reinforce expression of the whole pathway, making it independent of the external environmental cues which initiated it. This has been called *feedback fixation*, and it results in developmental determination in the classic embryological sense.

Many of the *Aspergillus* conidiation mutants are also defective in sexual reproduction. Thus, another conclusion to be drawn from these *A. nidulans* mutants is that there is some economy of usage of morphogenetic genes in different developmental processes. Different developmental modes presumably employ structural genes that are not uniquely developmental, but which function in numerous pathways, and have their developmental specificity bestowed upon them by the regulators to which they respond. This is epitomized in the idea that the key to eukaryote development is the ability to use relatively few regulatory genes to integrate the activities of many others.

Neurospora crassa forms two types of conidium: microconidia and macroconidia. *Microconidia* are small uninucleate spores, which are essentially fragmented hyphae. They are not well adapted to dispersal and are thought to serve primarily as "male gametes" in sexual reproduction. *Macroconidia* are more common and more abundant. They are large multinucleate, multicellular spores produced from aerial conidiophores. Conidiation (and sexual reproduction, too) in *N. crassa* seems to respond more to environmental signals than to complex genetic controls like those operating in *Aspergillus*. Macroconidia are formed in response to nutritional limitation, desiccation, change in atmospheric CO_2, and light exposure (blue light is most effective, and though light exposure is not essential, conidia develop faster and in greater numbers in illuminated cultures). In addition, a circadian rhythm provides a burst of sporulation each morning. When induced to form conidia, the *Neurospora* mycelium forms aerial branches, which grow away from the substratum and form many lateral branches that become conidiophores, which undergo apical budding to produce conidial chains.

The genetics of conidiation has been studied by means of mutation and molecular analysis. There are some parallels in terms of types of mutants obtained with *N. crassa* and *A. nidulans*, and a particular example would be the hydrophobic outer rodlet layer that is missing in the *N. crassa* "easily-wettable" (*eas*) and *A. nidulans rodA* mutants. Despite such functional analogies, there is no underlying similarity between the genetic architectures used by these two organisms to control conidiation. It is important that there is no evidence for regulatory genes in *N. crassa* similar to the *brlA-abaA-wetA* regulators of *A. nidulans*. Nevertheless, a large number of mutants have been isolated that have defects in particular stages of conidiation, although there is a general absence of analysis at the molecular level. Several conidiation (*con*) genes are known that encode transcripts that become more abundant at specific stages during conidiation. At least

four of these genes are expressed in all three sporulation pathways in *Neurospora* (macroconidia, microconidia, and ascospores), but others have specific localization to macroconidia. Many of the *con* genes, however, can be disrupted without affecting sporulation; thus, despite being highly expressed during sporulation, they presumably encode redundant or nonessential functions.

10.13 Sexual Reproductive Structures in Ascomycetes and Basidiomycetes

The conidiation mutants of *Aspergillus* and *Neurospora* make it clear that mycelium has a number of alternative developmental pathways open to it: continuation of hyphal growth, production of asexual spores, and progress into the sexual cycle. Sexual reproduction predominates over conidiation in many strains of *A. nidulans* collected from the wild when grown on normal media; laboratory strains carry a mutation (*veA* = velvet), which shifts the balance toward conidiation. Mutations to increased sexual reproduction at the expense of conidiation, however, were frequent among induced *A. nidulans* conidiation mutants. In contrast, some of the *Aspergillus* conidiation mutants (including *medA*, *stuA*, *yB*, and *acoA*) also exhibit defects in sexual reproduction, suggesting shared functions in the different morphogenetic pathways; unfortunately, far less attention has been given to sexual reproduction in these ascomycetes than to conidiation.

In Ascomycotina, the sexually produced asci are enclosed in an aggregation of hyphae termed an *ascoma*. Ascomata are not formed from hyphae that have taken part in the meiotic cycle; instead, they arise from nondikaryotic sterile hyphae that surround the ascogonial hyphae of the centrum. A variety of ascomata exist, including the open cuplike "discocarps" of *Peziza*, the flasklike perithecium (found, for example in *Neurospora* and *Sordaria*) and the completely closed cleistothecium formed by, for example, *Aspergillus*. Classical genetic approaches (e.g., identification of variant strains, application of complementation tests to establish functional cistrons, construction of heterokaryons to determine dominance/recessive and epistatic relationships and to indicate the sequence of gene expression) were used many years ago to establish a "developmental pathway" for perithecium formation in *Sordaria* (Fig. 10.7). This unfortunately remains the closest approach we have to a complete genetic pathway for ascome development.

Almost all of the research on sexual development in *Neurospora* has been aimed at understanding mating-type structure and function. Heterokaryons in which one nucleus carried a recessive color mutant have been used as genetic mosaics to show that perithecia of *Neurospora* arise from an initiating population of 100–300 nuclei, and that the perithecium wall is composed of three developmentally distinct layers. Twenty-nine comple-

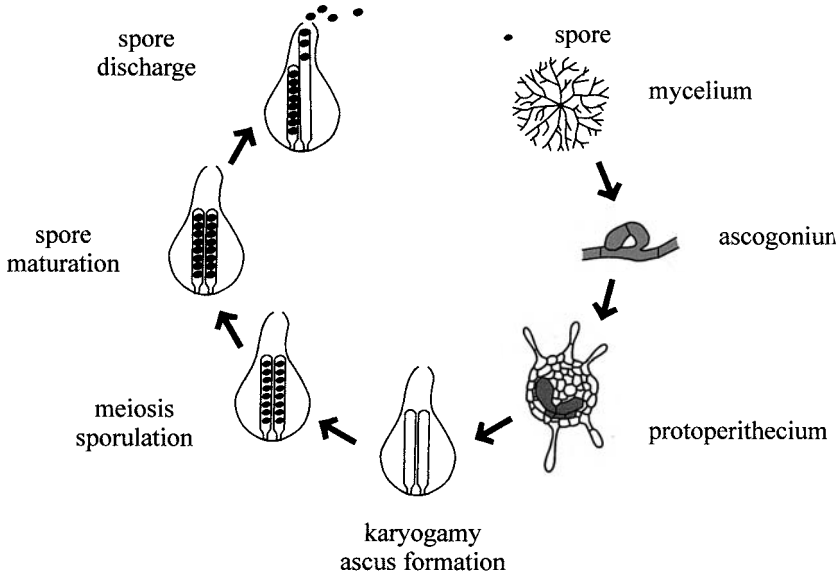


Fig. 10.7. Life-cycle diagram and perithecium developmental pathway of *Sordaria macrospora*. A variety of mutants are known that block the pathway at each of the stages represented by arrows, so the whole pathway is interpreted as being essentially a single sequence. Contrast this with the multiple parallel "subroutines" that characterize basidiomycete fruit body development (Fig. 10.8). (Figs. 10.5–10.8 based on Figs. 5.11–5.16 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

mentation groups (equivalent to functional genes) have been identified as being involved specifically in perithecium development. The *Escherichia coli* β-glucuronidase reporter gene has been used to study development of ascomata in *Pyrenopeziza brassicae*, and this work also revealed three tissue layers, but it also showed differential expression of the mating types. Both mating types are expressed in one of the layers, but the two mating types are expressed separately in each of the other two layers. The significance of extensive tissue layers in which only one mating type is expressed is unknown, but it may be analogous to differential expression of genes in dikaryotic hyphae of the basidiomycete *Schizophyllum commune*, which is thought to depend on change in proximity of nuclei carrying the mating-type factors.

Cleistothecium development in *Aspergillus nidulans* has been fully described, but the developmental observations have not been accompanied by extensive genetic analysis as yet. Apart from the involvement of conidiation mutants coincidentally noted earlier, a β-tubulin gene has been shown to be essential for sexual reproduction, and laccase activity is specifically located in cleistothecium primordia. Laccase enzymes have been associated

with several asexual and sexual reproductive processes in ascomycetes and basidiomycetes.

Increased phenoloxidase activity accompanies the initiation of fruit bodies in various basidiomycete species. The *cohesiveless* mutant of *Schizophyllum commune* has lost the ability to form hyphal aggregates and has no phenoloxidase activity in monokaryons. Dikaryons of *cohesiveless* could not form hyphal aggregates or fruiting bodies, but it regained phenoloxidase activity. Again, phenoloxidase is necessary but not sufficient for the formation of fruit body initials. When enzyme activity is measured in extracts of whole mycelium of *Coprinus congregatus* there is a simple correlation between the levels of laccase activity and the development of fruit body primordia, but localization studies reveal the true correlation to be with sensitivity to photoinduction. Light is able to induce primordia only in those regions of the mycelium that have high laccase activity. Perhaps laccase has a role in attaining developmental competence.

Several mutations have been described that can suppress or modify incompatibility reactions in *Podospora anserina*, and nearly all these modifying (*mod*) mutants affect protoperithecial formation and fertility. Several *mod* mutants have altered proteinase enzymes; others are thought to be affected in ribosomal structure and translation or to suffer from plasma membrane defects. Some of the mutants are defective in protoperithecial production and/or ascospore germination, but in others protoperithecial production is increased and occurs earlier than in the wild type. These observations suggest a model in which the key feature of fungal development is seen as the achievement of a quiescent state (= competence?), which is prerequisite for protoperithecial development. The quiescent state may have evolved as a survival mechanism under conditions of nutrient limitation. This interpretation equates production of the protoperithecium with production of a vegetative survival structure. It is interesting in view of the nonperithecial multicellular structures observed in *Sordaria* and the close genetic relationship that has been demonstrated between sclerotium production and the fruit body initiation pathway in the basidiomycete *Coprinus cinereus* (see later).

For the fruit bodies of basidiomycetes, which include mushrooms, toadstools, bracket fungi, puffballs, stinkhorns, and bird's nest fungi, the picture revealed by classical genetic approaches is less clear. One reason for this is that fruit bodies in basidiomycetes are normally formed by secondary mycelia, which are heterokaryotic. The co-existence of two (or more) nuclei, and, therefore, two or more genotypes, makes it difficult to study the genetics of development by conventional means. On the other hand, fruiting by monokaryotic mycelia has been recorded in many basidiomycetes, and these strains have allowed a start to be made on the genetic control of fruit body development. The frequency of monokaryons able to fruit differs drastically between genera: 27% of *Sistotrema* isolates form monokaryotic fruit bodies, 7% of *Schizophyllum* strains do so, but only 1 of 16 monokaryons

of *Coprinus cinereus*. Most "monokaryotic fruits" are abnormal structures that are usually incomplete or sterile, or both. This raises the question of whether genes that influence fruiting in monokaryons are relevant to the normal process of dikaryotic fruiting.

Several monokaryotic fruiting strains have been identified in the collection of *Polyporus ciliatus* isolates. Conventional genetic crosses made between them revealed three unlinked genes involved in monokaryotic fruiting: fi^+, which was thought to initiate monokaryotic fruiting; fb^+, which is seen as being responsible for "molding" the structure of the fruit initiated by fi^+ into a fruit body. The third gene, mod^+, appeared to direct development into a futile pathway leading to formation of nonfruiting mycelial masses called *stromata*. In the dikaryon mod^+ inhibited fruiting, but neither fi^+ nor fb^+ showed any expression even when homozygous. It is not known how these genes function. A very similar genetic system was found in analogous experiments with the agaric *Agrocybe aegerita*. Again, one gene, fi^+, was identified as being responsible for initiation of monokaryotic fruiting, and a second, fb^+, was considered to be responsible for modeling the initiated structures into fruit bodies. A contrast with the genes found in the polypore, *Polyporus*, was that the *Agrocybe* genes were found to influence fruiting in the dikaryon as well. Fertile fruit bodies were produced only by dikaryons carrying at least one allele of both fi^+ and fb^+.

Monokaryotic fruiting strains of *Schizophyllum commune* (called *hap*) show there is no correlation between monokaryotic and dikaryotic fruiting, and that monokaryotic fruiting is probably under polygenic control. Four genes of *S. commune* that control monokaryotic fruiting include two "fruiting initiation genes" (fi-1^+ and fi-2^+), either of which alone allow differentiation into fruit body initials of about 2 to 3 mm in size. When both are present, fruit body stems 6–8 mm long are formed. A third gene (fb^+) determines formation of complete monokaryotic fruit bodies. The fourth gene (st^+) prevents expression of the others. A monokaryon carrying st^+ produces only stromata and a homozygous st^+/st^+ dikaryon is also unable to fruit.

The other three genes have no affect on differentiation of fruit bodies in the dikaryon, but they do influence how quickly the dikaryon fruits. Dikaryons homozygous for all three monokaryotic fruiter genes fruit most rapidly. Dikaryons that do not carry any of the monokaryotic fruiter alleles fruit most slowly, but they *do* form fruiting bodies, which clearly implies a major difference in the impact that these genes have on the fruit body development pathway in the two types of mycelium. Increased frequency of fruiting in dikaryons made from monokaryotic fruiters has also been reported in *Lenzites trabea*. There are also several genes that enable monokaryons of *Schizophyllum commune* to initiate fruiting bodies in response to mechanical and chemical treatments; a total of eight genes have been identified, involved in four distinct pathways. These genes operate at a stage prior to the formation of aggregations of cells without defined shape

and may be distinct from those described earlier in this paragraph that produce structures with a recognizable stemlike shape.

The wide range of genetic factors involved in monokaryotic fruiting mirrors the range of physiological conditions that are able to promote such fruiting. Some of the genes identified in monokaryons do show expression in the dikaryon, but the role they might play is obscure. Another peculiarity is the induction of dikaryotic fruiting bodies on originally monokaryotic cultures of *Coprinus cinereus* when the latter are subjected to nutritional stress for several weeks to several months. In this case, nutritional stress may trigger a mating-type switch, which results in a conventional dikaryon being established. Spontaneous mating-type switching that enables homokaryotic mycelia to become fruiting dikaryons has occasionally been observed in *Agrocybe aegerita* and *Agaricus bitorquis*, but the molecular processes involved are unknown.

The only molecular observation made on fruit body induction is a DNA sequence that induced monokaryotic fruiting in strains of *Schizophyllum commune* into which it was introduced by transformation. The gene is called *FRT1*. Disruption of *FRT1* in the homokaryon results in a large increase in the expression of genes normally associated with enhanced growth of aerial mycelia, and with expression in the dikaryon. Disruption of *FRT1*, however, has no effect on either dikaryon growth or on the development of fruit bodies by the dikaryon. It is likely that *FRT1* normally acts as a repressor of dikaryon-expressed genes. The predicted sequence of the FRT1 protein suggests that it could be part of a signal transduction pathway involved in the regulation of genes that must be expressed in the dikaryon. The *pcc1* gene in *Coprinus cinereus* represents what might be a similar phenomenon. This recessive mutation arose in a homokaryotic strain that produced fertile fruit bodies after prolonged culture. Vegetative hyphae of the *pcc1* strain formed incomplete clamp connections (also called *pseudoclamps*; see Fig. 2.3), but *pcc1* is distinct from the *A* and *B* mating-type factors. Cloning and sequencing established that *pcc1* encodes a transcription factor that functions as part of the *A* mating-type factor pathway, and that the mutation in *pcc1* results in fruit body formation being released from repression in the homokaryon. Homologues of the *pcc1* sequence occur in the mating-type pathways of *Podospora anserina*, *Schizosaccharomyces pombe*, and *Ustilago maydis*.

The only organism in which any concerted attempt has been made to study the genetic control of fruit body formation by the dikaryon is *C. cinereus*. Dikaryons of *C. cinereus* can form sclerotia and basidiomata; monokaryons may also form sclerotia, but they normally do not form basidiomata. Initial steps in the development of both sclerotia and fruit bodies have been described separately and the descriptions are remarkably similar. In the formation of both structures, development from the mycelium involves similar patterns of hyphal aggregation, so the likeness observed

may indicate a shared initial pathway of development or coincidentally analogous separate, but parallel, pathways.

These possibilities were distinguished with the aid of monokaryons unable to form sclerotia, which is a phenotype that segregated in crosses as though controlled by a single major gene. Four *scl* (sclerotium-negative) genes were found; one, *scl*-4, caused abortion of developing fruit body primordia even when paired in the dikaryon with a wild-type nucleus, but the other *scl* genes behaved as recessive alleles in such heteroallelic dikaryons and were mapped to existing linkage groups. Homoallelic dikaryons (i.e., dikaryons in which both nuclei carried the same *scl* allele) were unable to form either sclerotia or fruit bodies. Because these single genetic defects blocked development of both dikaryon structures it was concluded that in the initial stages sclerotia and basidiomata share a common developmental pathway governed by the *scl* genes (Fig. 10.8). They are usually recessive when they mutate, so the pathway can proceed only in the heteroallelic dikaryon where the missing *scl* function is provided by the nucleus from the other parent.

The basic genetic control of dikaryon fruit body development in *Coprinus cinereus* has been examined by searching for developmental abnormalities among the survivors of mutagen-treated fragments of dikaryotic mycelium. Including spontaneous mutations, a total of 1594 were identified out of 10,641 dikaryotic survivors tested, and were classified into categories on the basis of the phenotype of the fruit body produced. The categories were: (1) *knotless*, no hyphal aggregations are formed; (2) *primordiumless*, aggregations are formed but they do not develop further; (3) *maturationless*, primordia are produced that fail to mature; (4) *elongationless*, stem fails to elongate, but cap development of the mushroom is normal (5) *expansionless*, stem elongation is normal, but the cap fails to open; (6) *sporeless*, few or no spores are formed in what may otherwise be a normal fruit body.

Because dikaryotic mutagen survivors were isolated, the genetic defects identified are all dominant. Elongationless mutants have been used to study stem elongation, and sporeless mutants have been used to study sporulation. These mutants suggest that fruit body development is organized into different pathways, which are genetically separate. Prevention of meiosis still permits the fruit body to develop normally, demonstrating, as do monokaryotic fruit bodies, that meiosis and spore formation are entirely separate from construction of the spore-bearing structure. It is also very significant that mutants were obtained with defects in either cap expansion or stem elongation. Both processes depend on enormous cell inflation in *Coprinus*, and the fact that they can be separated by mutation indicates that the same result (i.e., increase in cell volume) is achieved by different means.

There is a problem in accounting for the induction of dominant mutations at the high frequency observed in this study and the peculiarity that more than 72% of the mutants belong to just two phenotypes: there are

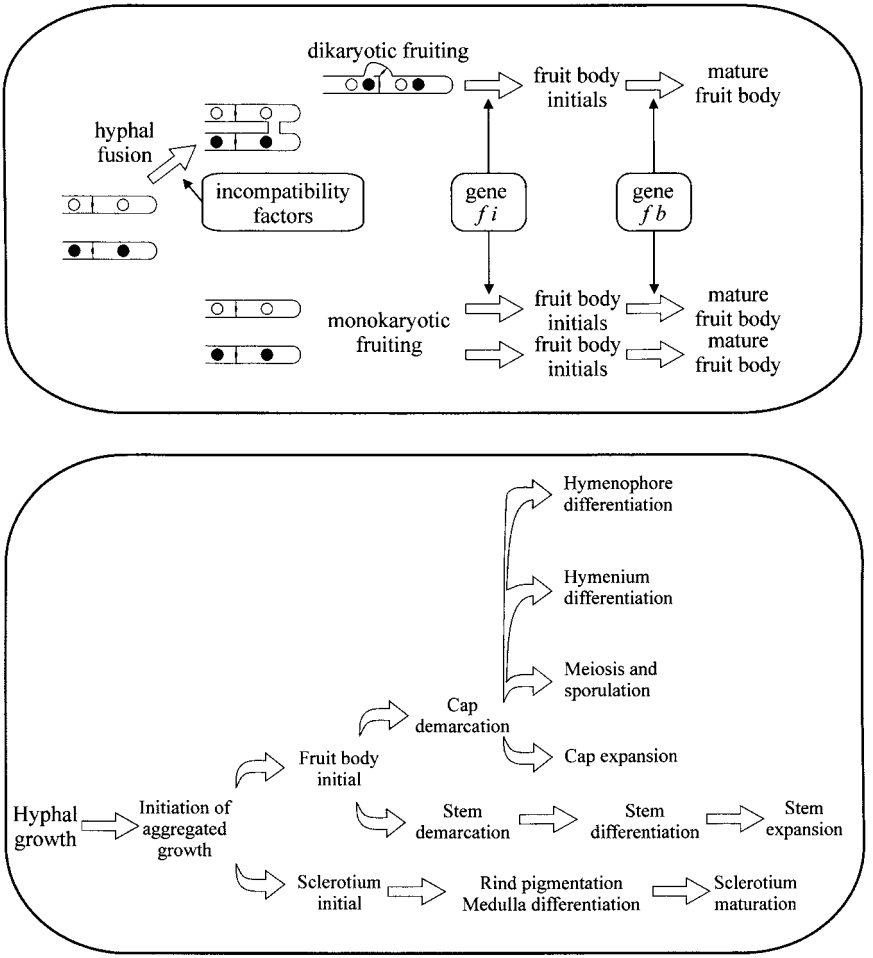


Fig. 10.8. Models for the genetic control of fruit body development in basidiomycetes. The top panel shows a proposed model for the action of major genes controlling mushroom formation in *Agrocybe aegerita*. The bottom panel shows the genetically distinct pathways involved in sclerotium and fruit body development in *Coprinus cinereus*. (Figs. 10.5–10.8 based on Figs. 5.11–5.16 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

595 maturationless and 582 sporeless isolates out of the total of 1582 induced mutants. These frequencies might suggest that genes involved in development may be easy to mutate, but an alternative interpretation is that the genes that were being caused to mutate were not those involved directly in development, but rather were genes that modify the dominance of pre-existing developmental variants. Dependence of dominance (or penetrance) on the modifying action of other genes is a well-established idea in

genetical theory, and could have considerable selective advantage in a system that imposes recessiveness on variants in genes concerned with development. The penetrance of *scl* genes in heteroallelic dikaryons of *C. cinereus* depends on the segregation of modifiers. Dominance modification has also been invoked to explain segregation patterns of a gene conferring resistance to *p*-fluorophenylalanine in *C. cinereus*. Because differentiation in basidiomycetes involves extensive protein processing, modifiers might be involved in processing signal sequences of structural proteins. In the presence of particular modifier alleles (i.e., those that cause the change in penetrance), signal processing might lead to normal structural proteins failing to reach their correct destination, or to abnormal proteins being partially corrected so that they do reach the target site, despite being defective.

Isolation of strains of *C. cinereus* with mutations in both mating-type factors (*Amut Bmut* strains) has opened up new possibilities for genetic analysis of morphogenesis in this organism. *Amut Bmut* strains are homokaryotic phenocopies of the dikaryon; that is, they emulate the dikaryon in that their hyphae have binucleate compartments and extend by conjugate nuclear division with the formation of clamp connections. In addition, the cultures can produce apparently normal fruit bodies. On the other hand, they are homokaryons, so they are able to produce asexual spores (usually called *oidia*) and, most importantly, contain only one (haploid) genetic complement. This last feature allows expression of recessive developmental mutations. These strains have been used in this way to study mutants in meiosis and spore formation, and in the formation of fruit body primordia. No overall fruit body developmental pathway has yet emerged, however, nor has any information about major regulators.

Genetic analysis of the sort discussed so far gives no guidance about the way in which genes causing developmental variants might exercise their effects. Among the first enzymes to be identified as having an important role in morphogenesis were glucanases involved in the degradation of fungal cell walls. The concept that cell wall materials are reutilized during morphogenesis originated with studies on *Schizophyllum commune*, and has received support from work with fruit bodies of *Flammulina velutipes* and *Coprinus congregatus* among basidiomycetes, as well as *Aspergillus nidulans* cleistothecia. The latter example is important because a mutant of *A. nidulans* that lacked α-1 $\rightarrow$ 3 glucan is unable to form cleistothecia, and mutants deficient in either cleistothecial formation or conidiation, or both, confirm there is at least a correlation between the presence of α-1 $\rightarrow$ 3 glucan, depletion of glucose, synthesis of α-1 $\rightarrow$ 3 glucanase, and cleistothecial formation.

Another important aspect of the sequence of studies on *A. nidulans* cleistothecium development is that it emphasizes the flexibility of the developmental process by showing that if glucan reserves are low, proteins may be utilized for cleistothecium formation. The exact nature of the nutrient limitation conditions determine whether glucans or proteins are used during

morphogenesis, but when circumstances demand, specific glucanase activity is replaced by specific proteinase action. This sort of flexible integration of enzyme activities to suit the prevailing conditions goes some way to explaining why only a small fraction of the genome is specific to morphogenesis, and why correspondingly few morphogenesis-specific polypeptides have been identified. A development-specific protein has been identified in sclerotia of *Sclerotinia sclerotiorum*, and a polypeptide specific to fruit body (ascomatal) development has been detected in *Neurospora tetrasperma* and localized to the mucilaginous matrix surrounding the asci and paraphyses. In *Sordaria brevicollis*, 17 out of more than 200 polypeptides detected after pulse labeling were found in perithecia. Only 15 polypeptides were found to be specifically expressed in fruit body primordia of *Schizophyllum commune*. Analysis of specifically transcribed RNA also suggest that expression of only a small proportion of the genome is devoted to morphogenesis in both *S. commune* and *Coprinus cinereus*. In the latter organism, only four so-called cap proteins have been found that were abundant in cap cells, but rare in the stem. Another example of differentially expressed proteins in *C. cinereus* is the appearance of lectin proteins, which specifically bind β-galactosides. Two of these galectins are expressed differently. The galectin known as *cgl1* is expressed in primordia and mature fruit bodies, whereas *cgl2* appears in the very earliest stages of fruit body initiation and is maintained until maturation. The function of these proteins is unknown, but they are excreted into the extracellular matrix and may be involved in cell–cell aggregation.

In situ hybridization has been used to demonstrate the reallocation of ribosomal-RNA between fruit bodies and their parental vegetative mycelium in *S. commune*; accumulation of fruiting-specific RNAs in the fruit body has also been demonstrated. Sequences cloned from among the fruiting-specific genes belong to a family that encode hydrophobins. These are cysteine-rich polypeptides that are excreted into the culture medium but polymerize on the wall of aerial hyphae as they emerge into the air (e.g., to form fruit body initials) and invest them with a hydrophobic coating. In *S. commune*, some hydrophobin genes are under control of the mating-type genes, and sequences coding for hydrophobins have been found in *Agaricus bisporus*, one of which specifically accumulates in the outer layers of mushroom caps (the "peel" tissue) during fruit body development. Hydrophobins, however, have been very widely encountered in fungi; about 20 have been recognized by gene sequence homology. They are small, secreted proteins comprised of 75–125 amino acids, with a high proportion of nonpolar amino acids, and eight cysteine residues spaced in a specific pattern ($X_{2–38}$-C-$X_{5–9}$-C-C-$X_{11–39}$-C-$X_{8–23}$-C-$X_{5–9}$-C-C-$X_{6–18}$-C-$X_{2–13}$, in which C = cysteine and X = any amino acid). Hydrophobin proteins have two domains, one hydrophilic and the other hydrophobic, and are capable of self-assembly at hydrophilic–hydrophobic interfaces (= interfacial self-assembly). They form amphipathic films that may be very insoluble. Protein

films formed by *S. commune* SC3 are insoluble in most aqueous and organic solvents.

The hydrophobins are a large and diverse family of proteins, which contribute to the nonspecific interactions that assist microorganisms to attach to surfaces. As such, they have been suggested to have roles in spore dispersal and adhesion (particularly in pathogens) as well as during morphogenesis. In the morphogenetic context it is important to remember that there are numerous hydrophobins that may function differently and at different times. The *S. commune* SC3 hydrophobin coats aerial hyphae and hyphae at the surface of fruit bodies; the SC4 hydrophobin coats voids (possibly air channels) within solid fruit body tissues and prevents them filling with fluid. Both confer hydrophobicity to these surfaces, but because hydrophobins form amphipathic layers, they can also make hydrophobic surfaces wettable. Teflon sheets immersed in SC3 hydrophobin become coated with a strongly adhering protein film that makes the plastic surface completely wettable. The hydrophobins alone suggest mechanisms that may be responsible for adherence of hyphae to each other and to other surfaces. They indicate more generally that the surface properties of the hypha can be controlled and manipulated to serve particular morphogenetic purposes as a result of specific gene expression.

Genes that encode hydrophobins expressed during the formation of emergent structures like aerial hyphae and fruit body initials are potential downstream targets of control genes involved in regulating fruit body development. Apart from the hydrophobins, relatively few such genes unfortunately have yet been identified. Some of the genes mentioned in the early paragraphs of this section are obvious candidates: genes that, when defective, affect fruit body initiation, the form and structure of the fruit body, meiosis, spore formation, and dispersal. None unfortunately have yet been cloned for molecular analysis, so the dynamics of their expression remain unknown. An exception might be the *ichijiku* (*ich1*) mutant of *Coprinus cinereus*. In the wild type of *C. cinereus*, a rudimentary fruit body cap can be clearly seen on the top of primordia, even those that are only about 1 mm in height. This rudimentary cap is missing in the *ich1* mutant, which arose as a spontaneous mutation in the progeny of a normal fruit body collected in the field. Because the cap is missing, there is no hymenophore, so the fruit body is sterile.

The mutant was called *ichijiku* because this means "fig" in Japanese; the Chinese characters mean "a fruit without flowering." The *ich1* gene product seems to be essential for cap formation, and in normal fruit body development the *ich1* transcript is specifically expressed in the cap. In addition, its abundance decreases as basidiospores are produced. The *ich1* gene encodes a large protein (1353 amino acids), the sequence of which contains nuclear targeting signals. This suggests that the Ich1 protein functions in the nucleus and may be a transcription regulator, although the sequence does not contain known DNA binding motifs. It is likely that Ich1 regulates the

expression of other genes required for cap and hymenophore development. The *ich1* mutant lacks the promoter region of the gene, and no *ich1* mRNA can be detected in the mutant. Other *hymenophoreless* mutants of *C. cinereus* have been isolated and have proved to be different from each other and different from *ich1*. As might be expected, lack of a hymenophore is a phenotype that can result from several genetic defects.

Whatever genes are directly involved in morphogenesis, they are presumably ultimately controlled in some way by the transcriptional regulators produced by the mating-type factors (Section 2.9). Most of the recognizable developmental-specific genes certainly seem to be transcriptionally regulated; however, the translational regulation observed in *Aspergillus* conidiation is a powerful means of relating entry into a developmental pathway to the nutritional status of the supporting mycelium. We will explain shortly how translational triggering might be more widely used as a regulatory strategy in higher fungi.

Another message, which comes clearly from these studies, is that recessive mutations can lead both to loss and gain of the ability to form multicellular structures. As examples we can cite the *scl* mutants of *C. cinereus*, which are involved in fruit body initiation and which have lost the ability to form sclerotia. Contrast these with the *fis* mutants, some of which cause monokaryotic fruiting, the *roc* gene, which causes stromatic proliferations of *C. cinereus*, and the *hap*, *fi*, and *fb* genes in *Schizophyllum*, which confer the ability on the monokaryon to form a fruit body, a phenotype that is normally a character of the dikaryon. Attempts have been made to simplify many of these observations into a single developmental pathway (Fig. 10.8), yet much of the evidence points to there being a number of discrete partial pathways that can run in parallel. This appears to be reflected in the fact that variation in fruit body morphology is common in higher fungi and can span generic and even wider taxonomic boundaries.

Consideration of these fruit body polymorphisms has led to the suggestion that normal morphogenesis may be an assemblage of distinct developmental subroutines. This concept views the genetic control of overall morphogenesis as being compartmentalized into distinct segments, which can be put into operation independently of one another. Thus, this model postulates subroutines for hymenophore, hymenium, stem, cap, and the like, which in normal development appear to be under separate genetic control (Fig. 10.8). In any one species they are thought to be invoked in a specific sequence that generates the particular ontogeny and morphology of that species, but the same subroutines may be invoked in a different sequence as an abnormality in that same species or as the norm in a morphologically different species. The model provides a unifying theme for categorizing fruit body ontogeny and for clarifying phylogenetic and taxonomic relationships.

Using what is known about the few systems that are reasonably well understood (mating-type factors and conidiation regulators), it is tempting to speculate on the genetic architecture that might underpin such a model.

A word of warning is necessary, though, because although there is a good catalogue of major similarities between fungi and other eukaryotes, there seem to be some major differences in gene regulation between the different groups of fungi. Efforts to express genes of filamentous fungi introduced into yeast have failed, and expression of ascomycetous genes in basidiomycetes has resulted, in most cases, in partial or total loss of regulation. Such observations imply that gene regulation mechanisms may be specific at a high taxonomic level that is certainly beyond the family level and perhaps at phylum or subphylum level. Although this is of great interest from the point of view of evolution, where does this leave attempts to use observations made in one group for prediction and speculation in other groups?

It probably has little effect on such speculations, as long as they do not attempt to explain the unknown in too much detail. The strategy of the regulation may be more similar than the tactics employed. Although the genetic structure may be different, there are many similarities at the level of functional expression. For example, mating-type factors seem to serve the same function in much the same way in most fungi in which they occur, despite their different gene structures. There is also an overriding impression that the membrane and hyphal surfaces are crucial players in morphogenesis. Hydrophobins are now known to be an extremely common feature throughout the fungi and represent the sorts of proteins that can manipulate the surface properties of hyphae. There must be many more such proteins awaiting discovery.

10.14 Genetic Control of Morphogenesis of Fungal Fruit Bodies

The keys to form and structure in fungi can be interpreted from the discussion so far and presented as a set of plausible mechanisms for the control of fungal morphogenesis. Key words at each stage of development in fungi seem to be competence, induction, and change. Competence is repeatedly encountered. Hyphae must be able to initiate the next step, but the next step is not inevitable. Competence may be genetic (e.g., mating types), but it is primarily a physiological state. Induction is the process by which the competent tissue is exposed to conditions that overcome some block to progress and allow the next stage to proceed. Change occurs when the competent tissue is induced. The next stage always involves a change in hyphal behavior and physiology. The change is usually quite drastic and represents an additional property to those already expressed (i.e., each developmental step takes the tissue to a higher order of differentiation).

Differentiated hyphal cells require reinforcement of their differentiation "instructions." This reinforcement is part of the context within which they normally develop (i.e., it is part of their network), but when removed from

their normal environment most differentiated hyphae revert to the mode of differentiation that characterizes vegetative hyphae. Hyphal differentiation is consequently an unbalanced process in comparison with vegetative hyphal growth. In most hyphal differentiation pathways the balance must be tipped in the direction of "differentiation" by the *local* microenvironment, which is, presumably, mainly defined by the local population of hyphae.

Another common feature is that morphogenesis is compartmentalized into a collection of distinct developmental processes (called *subroutines*). These separate (or parallel) subroutines can be recognized at the levels of organs (e.g., cap, stem, veil), tissues (e.g., hymenophore, context, pileipellis), cells (e.g., basidium, paraphysis, cystidium), and cellular components (e.g., uniform wall growth, growth in girth, growth in length, growth in wall thickness). They are distinct genetically and physiologically and may run in parallel or in sequence. When played out in their correct arrangement the morphology that is normal to the organism under consideration results. If some of the subroutines are disabled (genetically or through physiological stress), the rest may still proceed. This partial execution of developmental subroutines produces an abnormal morphology. Homologous subroutines can be recognized in different fungi, and gross differences in morphology can then be related to the different ways in which homologous subroutines are executed. The flowchart in Fig. 10.9 summarizes these notions.

The flexibility in the expression of developmental subroutines allows the fruit body to react to adverse conditions and still produce a crop of spores. It also illustrates that tolerance of imprecision is an important attribute of fungal morphogenesis. The ultimate flexibility, of course, is that the differentiation process can be abandoned in favor of vegetative hyphal growth and a reversion to the invasive mycelium. A lesser level of flexibility may be that an incompletely adapted cell type carries out a particular function.

There is no shortage of candidates when it comes to searching for mechanisms that might control fungal morphogenesis. Homologues and analogues of all of the mechanisms known in animals and plants can be found in fungi. For control at the genetic level the mating-type factors (Chapter 2) provide prime examples of transcriptional control elements able to regulate specific morphogenetic subroutines. The regulation involves transcriptional activation and repression, and further "complication" can be introduced, if necessary, by using intrachromosomal recombination to interchange regulatory cassettes.

Given the prevalence of data that indicate that hyphal systems (1) need to develop a state of competence before they are able to undertake a developmental pathway, and (2) can be precipitated into embarking upon a particular morphogenesis by a variety of environmental signals, it is difficult to believe that translational triggering and feedback fixation are not widely used as regulators throughout the higher fungi. Translational triggering is a mechanism that can relate a morphogenetic pathway to the development

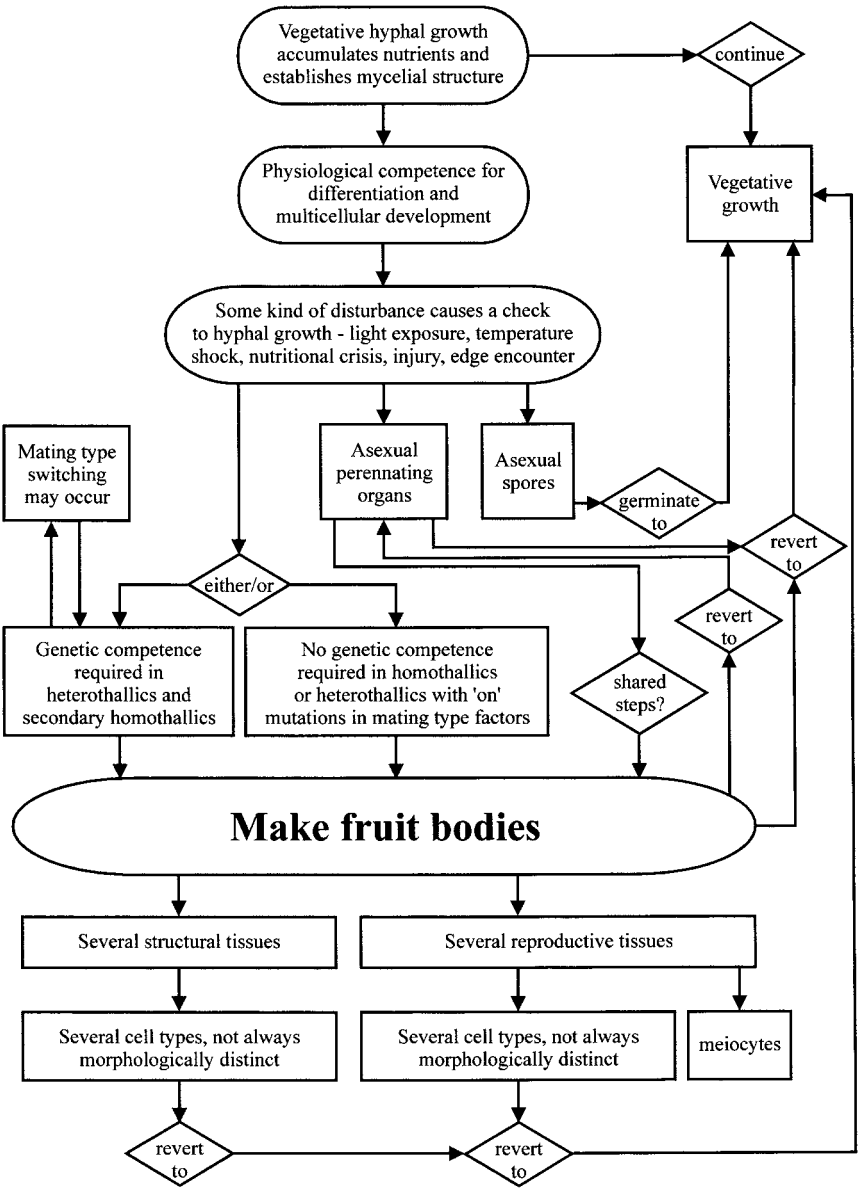


Fig. 10.9. Flowchart showing a simplified view of the processes involved in development of fruit bodies and other multicellular structures in fungi. (Figs. 10.9–10.11 based on Figs. 7.1–7.3 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

of competence on the one hand, and to initiation in response to environmental cue(s) on the other hand. There are indications from a wide range of physiological studies that nitrogen metabolism may be crucial in regulating morphogenesis. There would certainly be scope for associating particular differentiation pathways with particular aspects of metabolism; therefore, supply of specific aminoacyl-tRNA molecules might regulate entry into differentiation pathways by affecting translation of a controlling reading frame (trigger-ORF in Fig. 10.10).

The mechanism envisaged is in many ways similar to the attenuation mechanism that regulates several biosynthetic operons in bacteria. Because translation and transcription are so closely coupled in prokaryotes, attenuation regulates transcription. In an operon subject to attenuation, translation of mRNA commences soon after transcription begins. The RNA encodes a short (approximately 15 amino acid) leader peptide, which contains several adjacent codons for the amino acid product of the operon. When product levels are low, the corresponding aminoacyl-tRNA is limiting and the ribosome stalls at those codons. This allows a secondary structure to form in the mRNA that allows RNA polymerase to continue transcription of the structural genes of the operon. When the product of the operon is readily available, however, translation of the leader proceeds normally and an alternative secondary RNA structure allowing termination of transcription is formed. Attenuation provides a link between cellular levels of the product, which an operon is responsible for synthesizing, and transcription of the operon.

Attenuation depends on transcription and translation occurring simultaneously in time and space as they do in prokaryotes. Attenuation cannot operate in this way in eukaryotes because transcription and translation occur in different places and at different times. Nevertheless, there are several posttranscriptional stages (Section 10.9) at which a similar mechanism *could* regulate translation of a messenger transcript that coded for several reading frames. If the trigger-ORF contained adjacent attenuating codons for aminoacyl-tRNAs subject to variation in supply, stalling–nonstalling of translation of the trigger-ORF might determine whether the messenger transcript forms secondary structures that permit–do not permit translation of downstream reading frames. Note that either one or both components of the aminoacyl-tRNA may be the limiting factor, and the limitation may be imposed by compartmentalization. That is, amino acid or a specific tRNA (or, presumably, an aminoacyl-tRNA synthetase) may be compartmentalized, regulated in local concentration, or both.

The interpretation offers a way by which a competent tissue can be released to undertake differentiation by a range of physiological events. Competence is interpreted to mean that messenger transcripts for the necessary regulators (and perhaps some key structural genes) are produced but not fully translated because an upstream sequence (trigger-ORF) prevents translation. There may be a number of different such transcripts with

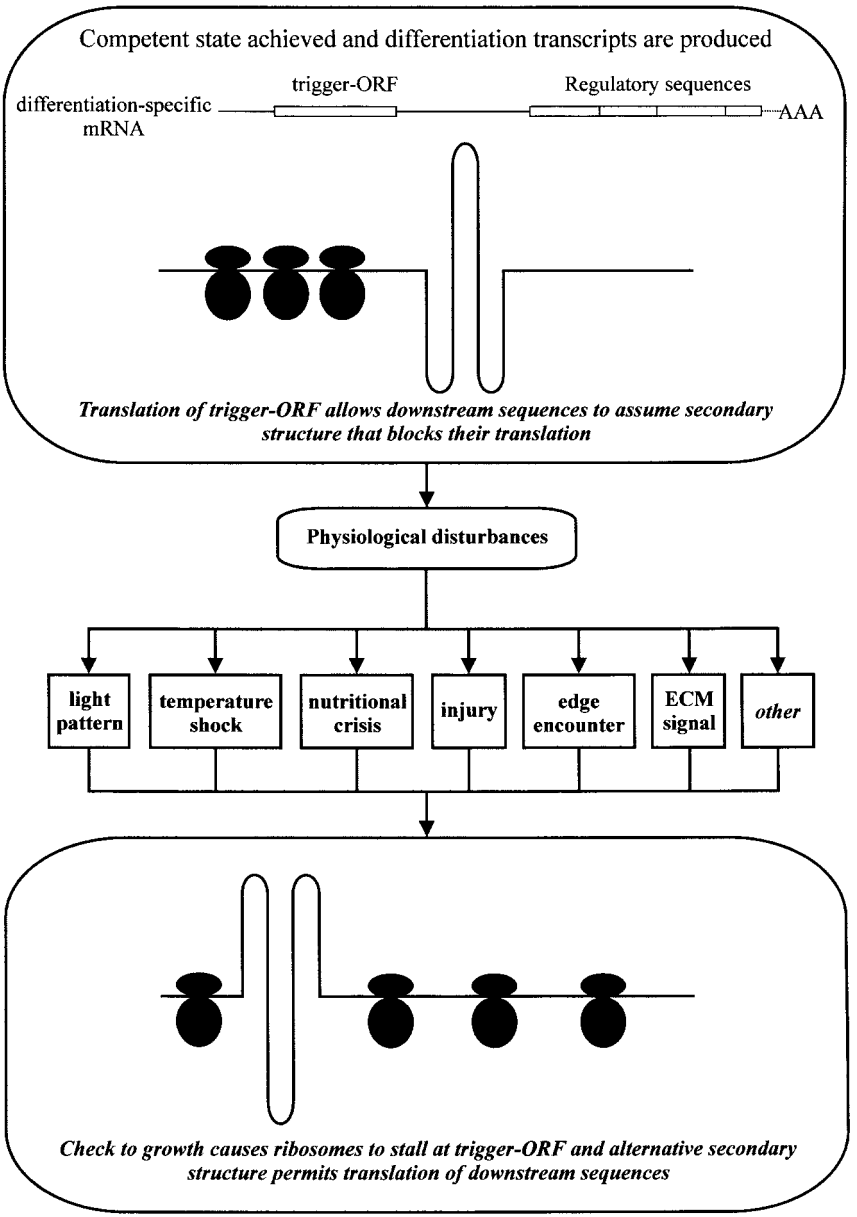


Fig. 10.10. Translational triggering adopted as a general model for entry of competent tissues into fungal pathways of differentiation. (Figs. 10.9–10.11 based on Figs. 7.1–7.3 in Moore (1998), Fungal Morphogenesis, Cambridge University Press.)

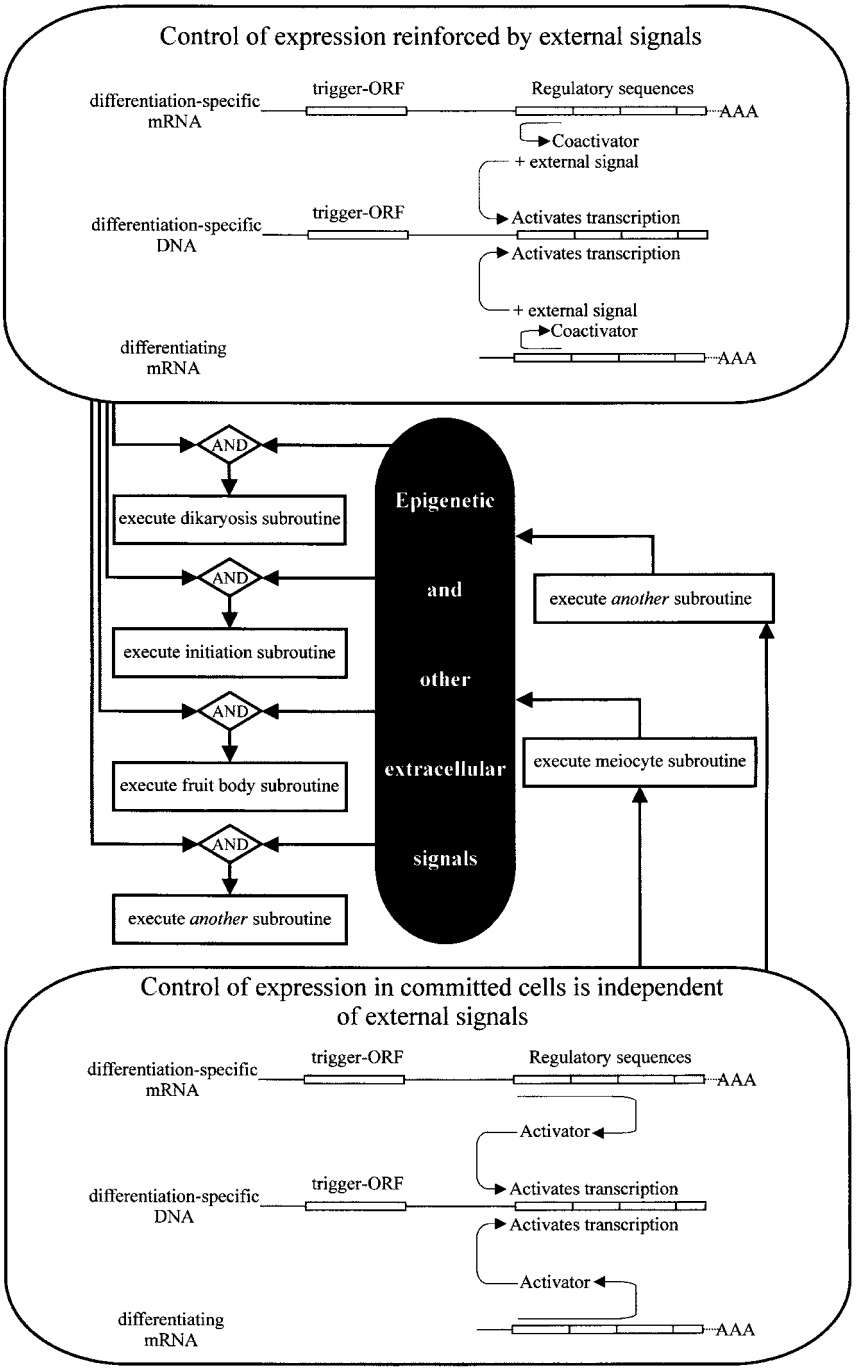
regulators that correspond to the different pathways upon which the competent cell can embark, their trigger-ORFs responding to separate physiological events (Fig. 10.10). On the other hand, there may be a number of similar transcripts in different cellular compartments so that the translational trigger can be released by the particular activities of those compartments with the result that one differentiation process may be triggered by different physiological events. It could also be that such a transcript is limited to one compartment, and even to one type of vesicle, perhaps, from which the trigger molecule can be excluded until some highly specific and/or localized physiological change occurs.

There is unfortunately no direct evidence for any of these speculations, although a variety of physiological signals and stresses cause translation-level controls to direct competent fungal tissues to undertake specific differentiation processes. A comparison with the operation of mating-type factors makes it reasonable to suggest that the translational trigger could immediately lead to translation of components of highly specific transcription activators and inhibitors, which then regulate gene sequences required for the differentiation that has been initiated. These, or their eventual products, may be involved in feedback fixation of the differentiation pathway.

Feedback fixation is the outcome of feedback activation and autoregulation that together reinforce expression of the whole regulatory pathway to make it independent of the external environmental cues that initiated it. Feedback fixation results in developmental determination in the classic embryological sense. The epigenetic aspects of the network governing fungal morphogenesis start with feedback fixation, but also include signals from outside the cell (Fig. 10.11). The fungal extracellular matrix is extensive and complex. Its reaction to, and interaction with, the environment can be communicated to the intracellular environment to modify cytoplasmic activity. Because neighboring cells are components of the external environment, it must be the case that the activity of one hyphal cell is modulated by changes made to the extracellular matrix by a neighboring hyphal cell.

On this interpretation, therefore, continued progress in differentiation for most fungal cells requires continued reinforcement from their local microenvironment. This may involve production of location- and/or time-specific extracellular matrix molecules, or any of a range of smaller molecules, that might be classed as hormones or growth factors. Smaller molecules might exert their effects by being taken up into the cell. Uptake, however, is not necessary. Any of these molecules may also affect relations between integrins and the existing extracellular matrix. As a result there could be direct effects on the cytoskeleton, which are able to cause immediate metabolic changes in one or more cellular compartments, or to influence gene transcription directly.

Connections to the extracellular matrix may also be involved in that other great enigma: the control of hyphal branching. By varying


Control of expression reinforced by external signals
differentiation-specific mRNA
trigger-ORF
Regulatory sequences
AAA
Coactivator
+ external signal
differentiation-specific DNA
trigger-ORF
Activates transcription
Activates transcription
+ external signal
Coactivator
differentiating mRNA
AAA
AND
execute dikaryosis subroutine
AND
execute initiation subroutine
AND
execute fruit body subroutine
AND
execute another subroutine
Epigenetic and other extracellular signals
execute another subroutine
execute meiocyte subroutine
Control of expression in committed cells is independent of external signals
differentiation-specific mRNA
trigger-ORF
Regulatory sequences
AAA
Activator
differentiation-specific DNA
trigger-ORF
Activates transcription
Activates transcription
Activator
differentiating mRNA
AAA

extracellular matrix–membrane or wall–membrane connections, external signals may be able to specify branch initiation sites. Internal cytoskeletal architecture could similarly also arrange specific membrane–wall connections to become branch initiation sites. Branch initiation sites specified in these ways may then become gathering sites for the molecules that create a new hyphal tip. The branch would consequently emerge in a position precisely defined by the stimulation of generalized cytoskeleton–membrane–wall connections by a positional stimulus. The focus of these hypothetical regulatory activities is, obviously, the hyphal wall, its surface, and the immediate extracellular environment. These are features about which we are very ignorant and urgent and extensive research on these topics is necessary. The key to fungal morphogenesis lies in understanding how that which is outside a hypha can influence that which goes on inside the hypha in a time- and place-dependent manner. We are still a long way from reaching that understanding.

Publications and Websites Worth a Visit

Banuett, F. (1998). Signaling in the yeasts: an informational cascade with links to the filamentous fungi. *Microbiology and Molecular Biology Reviews* **62**, 249–274.

Belmont, A.S., Dietzel, S., Nye, A.C., Strukov, Y.G. & Tumbar, T. (1999). Large-scale chromatin structure and function. *Current Opinion in Cell Biology* **11**, 307–311.

Biological Procedures Online, a library of step-by-step recipes at http://www.sciencemedcentral.com.

Bölker, M. (1998). Sex and crime: heterotrimeric G proteins in fungal mating and pathogenesis. *Fungal Genetics and Biology* **25**, 143–156.

Chiu, S.W. & Moore, D. (1990). A mechanism for gill pattern formation in *Coprinus cinereus*. *Mycological Research* **94**, 320–326.

Cogoni, C. & Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. *Current Opinion in Genetics & Development* **10**, 638–643.

De Groot, P.W.J., Schaap, P.J., Sonnenberg, A.S.M., Visser, J. & Van Griensven, L.J.L.D. (1996). The *Agaricus bisporus hypA* gene encodes a hydrophobin and specifically accumulates in peel tissue of mushroom caps during fruit body development. *Journal of Molecular Biology* **257**, 1008–1018.

Fig. 10.11. Feedback fixation adopted as a general model for maintaining progress through fungal pathways of differentiation. In this flow chart the box at the top shows the type of feedback fixation process envisaged to apply to most developmental subroutines in which epigenetic reinforcement from the local microenvironment is needed to interact with coactivators in order to maintain the feedback activation loop. In the bottom panel, the alternative of direct feedback fixation independently of other signals is shown as being applicable to cell types that show developmental commitment; only meiocytes are known to be committed, but there may be other committed cell types.

De Groot, P.W.J., Schaap, P.J., Van Griensven, L.J.L.D. & Visser, J. (1997). Isolation of developmentally regulated genes from the edible mushroom *Agaricus bisporus*. *Microbiology* **143**, 1993–2001.

Developmental Biology Virtual Library is at http://sdb.bio.purdue.edu/other/vl_db.html.

Developmental Biology Cinema is at http://sdb.bio.purdue.edu/dbcinema.

Dreyfuss, G. & Struhl, K. (1999). Nucleus and gene expression. Multiprotein complexes, mechanistic connections and nuclear organization. *Current Opinion in Cell Biology* **11**, 303–306.

Finnegan, E.J., Peacock, W.J. & Dennis, E.S. (2000). DNA methylation, a key regulator of plant development and other processes. *Current Opinion in Genetics & Development* **10**, 217–223.

Gancedo, J.M. (2001). Control of pseudohyphae formation in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* **25**, 107–123.

Grosveld, F. (1999). Activation by locus control regions? *Current Opinion in Genetics & Development* **9**, 152–157.

Habu, Y., Kakutani, T. & Paszkowski, J. (2001). Epigenetic developmental mechanisms in plants: molecules and targets of plant epigenetic regulation. *Current Opinion in Genetics & Development* **11**, 215–220.

Hicke, L. (1999). Gettin' down with ubiquitin: turning off cell-surface receptors, transporters and channels. *Trends in Cell Biology* **9**, 107–112.

Horton, J.S., Palmer, G.E. & Smith, W.J. (1999). Regulation of dikaryon-expressed genes by *FRT1* in the basidiomycete *Schizophyllum commune*. *Fungal Genetics and Biology* **26**, 33–47.

Hsieh, C.-L. (2000). Dynamics of DNA methylation pattern. *Current Opinion in Genetics & Development* **10**, 224–228.

Idekar, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R. & Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934.

Johnston, M. (1999). Feasting, fasting and fermenting: glucose sensing in yeast and other cells. *Trends in Genetics* **15**, 29–33.

Jones, P.A. (1999). The DNA methylation paradox. *Trends in Genetics* **15**, 34–37.

Kass, S.U., Pruss, D. & Wolffe, A.P. (1997). How does DNA methylation repress transcription? *Trends in Genetics* **13**, 444–449.

Kornberg, R.D. (1999). Eukaryotic transcription control. *Trends in Genetics* **15**, M46–M49.

Kron, S.J. & Gow, N.A.R. (1995). Budding yeast morphogenesis: signaling, cytoskeleton and cell cycle. *Current Opinion in Cell Biology* **7**, 845–855.

Kües, U. (2000). Life history and developmental processes in the basidiomycete *Coprinus cinereus*. *Microbiology and Molecular Biology Reviews* **64**, 316–353.

Kües, U. & Liu, Y. (2000). Fruiting body production in basidiomycetes. *Applied Microbiology & Biotechnology* **54**, 141–152.

Legrain, P., Wojcik, J. & Gauthier, J.-M. (2001). Protein–protein interaction maps: a lead towards cellular functions. *Trends in Genetics* **17**, 346–352.

Lipshitz, H.D. & Smibert, C.A. (2000). Mechanisms of RNA localization and translational regulation. *Current Opinion in Genetics & Development* **10**, 476–488.

Madhani, H.D. & Fink, G.R. (1998). The control of filamentous differentiation and virulence in fungi. *Trends in Cell Biology* **8**, 348–353.

Malik, S. & Roeder, R.G. (2000). Transcriptional regulation through Mediator-like coactivators in yeast and metazoan cells. *Trends in Biochemical Sciences* **25**, 277–283.

Martienssen, R.A. & Richards, E.J. (1995). DNA methylation in eukaryotes. *Current Opinion in Genetics & Development* **5**, 234–242.

Matzke, M.A., Matzke, A.J.M., Pruss, G.J. & Vance, V.B. (2001). RNA-based silencing strategies in plants. *Current Opinion in Genetics & Development* **11**, 221–227.

Melcher, K. (1997). Galactose metabolism in *Saccharomyces cerevisiae*: a paradigm for eukaryotic gene regulation. In *Yeast Sugar Metabolism* (F.K. Zimmermann & K.-D. Entian, eds.), pp. 235–269. Technomic Publishing Co., Inc.: Lancaster, PA.

Miller, B.L. (1993). Brushing-up on bristles—complex genes and morphogenesis in molds. *Trends in Genetics* **9**, 293–295.

Moore, D. (1998). *Fungal Morphogenesis*. Cambridge University Press: New York.

Muraguchi, H. & Kamada, T. (1998). The *ich1* gene of the mushroom *Coprinus cinereus* is essential for pileus formation in fruiting. *Development* **125**, 3133–3141.

Murata, Y., Fujii, M., Zolan, M.E. & Kamada, T. (1998). Molecular analysis of *pcc1*, a gene that leads to *A*-regulated sexual morphogenesis in *Coprinus cinereus*. *Genetics* **149**, 1753–1761.

Pérez-Martín, J. (1999). Chromatin and transcription in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* **23**, 503–523.

Preiss, T. & Hentze, M.W. (1999). From factors to mechanisms: translation and translational control in eukaryotes. *Current Opinion in Genetics & Development* **9**, 515–521.

Sánchez-Martínez, C. & Pérez-Martín, J. (2001). Dimorphism in fungal pathogens: *Candida albicans* and *Ustilago maydis*—similar inputs, different outputs. *Current Opinion in Microbiology* **4**, 214–221.

Selker, E.U. (1997). Epigenetic phenomena in filamentous fungi: useful paradigms or repeat-induced confusion? *Trends in Genetics* **13**, 296–301.

Timberlake, W.E. (1993). Translational triggering and feedback fixation in the control of fungal development. *Plant Cell* **5**, 1453–1460.

Tucker, C.L., Gera, J.F. & Uetz, P. (2001). Towards an understanding of complex protein networks. *Trends in Cell Biology* **11**, 102–106.

Van Griensven, L.J.L.D. (2000). *Science and Cultivation of Edible Fungi, Mushroom Science XV*, vols. 1 & 2. Balkema: Rotterdam.

Vaucheret, H. & Fagard, M. (2001). Transcriptional gene silencing in plants: targets, inducers and regulators. *Trends in Genetics* **17**, 29–35.

Wessels, J.G.H. (2000). Hydrophobins, unique fungal proteins. *Mycologist* **14**, 153–159.

Wu, C.-T. & Morris, J.R. (1999). Transvection and other homology effects. *Current Opinion in Genetics & Development* **9**, 237–246.

Historical Publications Worth Knowing About

Buller, A.H.R. (1909). *Researches on Fungi*, vol. 1. Longman, Green & Co.: London.

Buller, A.H.R. (1922). *Researches on Fungi*, vol. 2. Longman Green & Co: London.

Buller, A.H.R. (1931). *Researches on Fungi*, vol. 4. Longmans, Green & Co.: London.

Chang, S.T. & Hayes, W.A. (1978). *Biology and Cultivation of Edible Mushrooms*. Academic Press: New York & London.

Chiu, S.W., Moore, D. & Chang, S.T. (1989). Basidiome polymorphism in *Volvariella bombycina*. *Mycological Research* **92**, 69–77.

Esser, K. & Meinhardt, F. (1977). A common genetic control of dikaryotic and monokaryotic fruiting in the basidiomycete *Agrocybe aegerita*. *Molecular & General Genetics* **155**, 113–115.

Esser, K., Saleh, F. & Meinhardt, F. (1979). Genetics of fruit body production in higher basidiomycetes. II. Monokaryotic and dikaryotic fruiting in *Schizophyllum commune*. *Current Genetics* **1**, 85–88.

Esser, K. & Straub, J. (1958). Genetische Untersuchungen an *Sordaria macrospora* Auersw., Kompensation und induktion bei genbedingten Entwicklungsdefekten. *Zeitschrift für Vererbungslehre* **89**, 729–746.

Flegg, P.B., Spencer D.M. & Wood, D.A. (1985). *The Biology and Technology of the Cultivated Mushroom*. John Wiley & Sons: New York.

Hoge, J.H.C., Springer, J. & Wessels, J.G.H. (1982). Changes in complex RNA during fruit-body initiation in the fungus *Schizophyllum commune*. *Experimental Mycology* **6**, 233–243.

Leslie, J.F. & Leonard, T.J. (1979). Three independent genetic systems that control initiation of a fungal fruiting body. *Molecular & General Genetics* **171**, 257–260.

Moore, D. (1981). Developmental genetics of *Coprinus cinereus*: genetic evidence that carpophores and sclerotia share a common pathway of initiation. *Current Genetics* **3**, 145–150.

Moore, D., Casselton, L.A., Wood, D.A. & Frankland, J.C. (1985). *Developmental Biology of Higher Fungi*. Cambridge University Press: Cambridge, U.K.

Moore, D., Elhiti, M.M.Y. & Butler, R.D. (1979). Morphogenesis of the carpophore of *Coprinus cinereus*. *New Phytologist* **83**, 695–722.

Mulder, G.H. & Wessels, J.G.H. (1986). Molecular cloning of RNAs differentially expressed in monokaryons and dikaryons of *Schizophyllum commune* in relation to fruiting. *Experimental Mycology* **10**, 214–227.

Watling, R. (1971). Polymorphism in *Psilocybe merdaria*. *New Phytologist* **70**, 307–326.

Yashar, B.M. & Pukkila, P.J. (1985). Changes in polyadenylated RNA sequences associated with fruiting body morphogenesis in *Coprinus cinereus*. *Transactions of the British Mycological Society* **84**, 215–226.

Ye, X.S., Wu, L. & Osmani, S.A. (2001). Signal transduction in filamentous fungi: regulation of protein kinases. In *Molecular and Cellular Biology of Filamentous Fungi* (N.J. Talbot, ed.), pp. 157–174. Oxford University Press: Oxford, U.K.

Index