Serge Morand
François Beaudeau
Jacques Cabaret *Editors*

New Frontiers of Molecular Epidemiology of Infectious Diseases

Springer

New Frontiers of Molecular Epidemiology of Infectious Diseases

Serge Morand • François Beaudeau
Jacques Cabaret
Editors

New Frontiers of Molecular Epidemiology of Infectious Diseases

Springer

Editors
Serge Morand
Institut des Sciences de l'Evolution
CNRS, IRD, Université Montpellier 2
Montpellier Cedex 05
France
serge.morand@univ-montp2.fr

François Beaudeau
Unit of Animal Health Management
Veterinary School-INRA, BP 40706
Nantes Cedex 03
France
francois.beaudeau@oniris-nantes.fr

Jacques Cabaret
INRA, BASE, Ecology
and Genetics of Parasitology
Nouzilly
France
jacques.cabaret@tours.inra.fr

ISBN 978-94-007-2113-5 e-ISBN 978-94-007-2114-2
DOI 10.1007/978-94-007-2114-2
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011935736

Foreword

Dismissing the frontiers of species, and of ecosystems, the pathogens become major threats to man, its livestock and wildlife animals, in a global world facing global changes. Thus, the emerging and re-emerging infectious diseases are one the main scientific challenges that agricultural and environmental research, like medical research, has to face, transcending the frontiers of disciplines, mandates and reserved domains of expertise.

The oft-repeated sanitary events of the last decades showed the tremendous capacity of adaptation and resilience of pathogens, how they can profit, from the diversity of their populations for the spreading dynamics and/or the diffusion of a part of their genome. Therefore, it is of major importance to bring to the health stakeholders, the up-to-date knowledge on molecular and modelling tools in epidemiology of infectious diseases, the best and new concepts in population genetics and evolutionary ecology. However, it is necessary to integrate all these new approaches to get a better overview of the host pathogens and/or their vectors interactions, the transmission dynamics and therefore to elaborate risk analysis models, finally enabling to put forward innovation for the prevention and the control of infectious diseases.

This book is the formal results of exchanges that occurred during a Research school organised by the Animal Health Division of INRA in 2008. The interactions between participants coming from very varied backgrounds being so rich that one wanted to capitalize on its outputs in terms of new research avenues, by editing a book for students and researchers in public and/or veterinary health, in epidemiology, in health of natural fauna. The different chapters reflect in parts the sessions that occurred during this very fruitful training course on molecular epidemiology and population genetics, finally leading to one could call the integrated concept of evolutionary epidemiology. Several authors that did not participate to this Research school also contributed to this book, which is illustrated by a variety of examples, turning it into an essential review for the scientists interested in the so large and unrestrained domain of the epidemiology of infectious diseases.

It is a great pleasure for me to acknowledge the editors and colleagues, J Cabaret, S Morand and F Beaudeau, and all the contributors, for this book crossing the frontiers of disciplines, in a major domain of the agricultural and environmental research.

Dr Gilles Aumont
Formerly Head of Animal Health Division
Head of International Scientific Partnerships
French National Institute for Agricultural Research (INRA)

Contents

Part I
Methods and Tools in Molecular Epidemiology

Chapter 1
Introduction

Jacques Cabaret, Serge Morand, and François Beaudeau

Molecular epidemiology can be defined as a focus on the contribution of potential genetic and environmental risk factors, identified at the molecular level, to the etiology, geographic distribution and prevention of disease within and across populations. It emerged from the integration of molecular biology into traditional epidemiologic research. Molecular epidemiology has recently broaden its focuses due to the development of molecular tools but also by incorporating advances of other fields such as mathematical epidemiology, molecular ecology, population genetics and evolution. Facing new risks of emerging and re-emerging infectious diseases that are threats for humans, their livestock, and wildlife animals, the objectives of molecular epidemiology include:

– the development of molecular tools, genotyping and gene expression
– the incorporation of concepts and results of population genetics of infectious diseases
– the integration of recent advances in theoretical epidemiology, modelling, and evolutionary ecology of diseases
– a better understanding of transmission for the development of risk factors analyses.

J. Cabaret (✉)
INRA, UR1282 Infectiologie Animale et Santé Publique, F-37380 Nouzilly, France
e-mail: jacques.cabaret@tours.inra.fr

S. Morand
Institut des Sciences de l'Evolution, CNRS-IRD-UM2, CC65,
Université Montpellier 2, F-34095 Montpellier, France

Cirad AGIRs, F-34398 Montpellier, France
e-mail: serge.morand@univ-montp2.fr

F. Beaudeau
INRA, Oniris, UMR1300 Bio-agression, épidémiologie et analyse de risque en santé animale,
BP40706, 44307 Nantes, France
e-mail: francois.beaudeau@oniris-nantes.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_1,

Thirty-three specialists contribute 13 chapters to this volume exploring the wide range of approaches being used in the field of molecular epidemiology. Coverage includes: (i) methods and tools in molecular epidemiology, (ii) contribution of population genetics, (iii) contribution of population and evolutionary epidemiology, (iv) case studies where we may appreciate the interactions of new tools, population genetics and phylogeny, and (vi) new integrations regarding the evolution of pathogens facing heavy selective pressure from drugs, their largely increased dispersal due to the travel facilities, and finally the framework for understanding new emergence process of diseases.

Methods and tools for epidemiology have enormously increased and are still progressing in refinement. Instead of covering all the new methods available in epidemiology of the various pathogens, which could be the subject for another whole book, we concentrated on bacterial pathogens in the unique chapter of methods and tools, which intends to show the array of possibilities and the interest of characterizing pathogens. This chapter entitled "Molecular typing of bacterial pathogens: A tool for the epidemiological study and control of infectious diseases" by Marie Hallin, Ariane Deplano and Marc J. Struelens, demonstrates that nowadays molecular typing is an integral part of the public health microbiology toolbox. Typing is used both to investigate outbreaks and enhance the resolution of disease surveillance at different community levels: (i) locally, by clinical or public health laboratories; (ii) nationally, by reference laboratories or (iii) globally, through international surveillance networks. The authors provide an overview of currently available and emerging technologies for typing human bacterial pathogens, discusses their suitability to different levels of use and reviews examples of integrated typing in advanced surveillance systems. It is already the case in viruses like the human influenza where typing has been for long a tool for decision in vaccine construction and it will probably be of great help for protozoa and macroparasites.

The contribution of population and evolutionary genetics to molecular epidemiology has not yet been really investigated. The chapter on "Molecular epidemiology and evolutionary genetics of infectious diseases" by Michel Tibayrenc, argues why molecular epidemiology should not be conceived without an extensive use of the concepts of population genetics and evolutionary biology. Moreover, it stresses that characterizing pathogens should open up to evaluating the impact of pathogens' genetic diversity on their relevant medical properties (downstream studies). This chapter provides insights on the interest to link epidemiology of diseases to population genetics and evolution and offers a modern and easy to use framework for classical epidemiologists that may be lost in the forest of concepts expressed in the area. The chapter on "Population genetics and molecular epidemiology of infectious diseases" by Christine Chevillon, Thierry de Meeûs, Karen D. McCoy is dedicated to the importance of population genetics. From population genetics' analyses, inferences can be made on the reproductive modes of the species in question (hence, on patterns of genetic transmission along successive generations), as well as on the demographic functioning of the studied populations (i.e., population sizes, dispersal rates among populations, etc.). The authors present the main bases of population genetics theory and illustrate its interest for epidemiological issues via different case studies on parasite species.

The contribution of population and evolutionary epidemiology is illustrated by modelling infection and evolution of virulence. Modelling infection by Pauline Ezanno, Elisabeta Vergu, Michel Langlais, and Emmanuelle Gilot-Fromont shows that mathematical modelling is a valuable tool to analyze transmission and disease evolution, to test assumptions and analyse scenarios, and to predict outcomes of the host-pathogen interactions. This chapter aims at introducing basic concepts and methods of epidemiological modelling. They present the main principles of model building and analysis, using simple biological and also mathematical systems. They also provide an overview of the methods that can be employed to describe more complex systems. They finally illustrate how the modelling approach may help for evaluation of control strategies. The chapter on "Evolution of virulence: Intuitions and models" by Jean-Baptiste Ferdy and Sylvain Gandon reviews some of the classic theoretical results on virulence evolution in pathogens. Their goal is to contrast an intuitive vision on evolution of virulence to results that can be obtained using simple modeling tools. The main conclusion is that natural selection does not maximize the basic reproductive rate R0. This is because R0 quantifies invasion potential in a context where most hosts are susceptible, while selection in endemic situations acts in circumstances where many hosts are infected. They conclude on how the use of simple models can shed light on some experiments and, in return, how some experiment data is required to improve our theoretical predictions on pathogen evolution.

Three case studies are presented in order to illustrate how in real life, molecular epidemiology may play an important role, from diagnostic to control. One chapter by Philippe Lanotte reports on tuberculosis that remains a major public health problem. The development of molecular biology since early 1990s and recent advances according to genomics offer new opportunities to understand the epidemiological dissemination of strains from patient scale to the world scale. Molecular methods were initially developed to confirm genetic link between *M. tuberculosis* strains isolated in similar epidemiological circumstances such as intra familial transmission, nosocomial transmission, distinction between exogenous re-infection or relapse and to explore suspected transmission chain. The molecular methods based on mobile genetic elements, especially insertion sequences, or repetitive DNA sequences showed limits and were supplanted by Single Nucleotide Polymorphisms (SNP), Large Sequence Polymorphism (LSP) also called Regions of Differences (RD) that could help to trace the origin of infection. The second case by Magali Chabé, Jean-Pierre Hugot, and Eduardo Dei-Cas studies the *Pneumocystis* molecular phylogeny: a way to understand both pneumocystosis natural history and host taxonomy. The genus Pneumocystis comprises numerous fungal pathogens dwelling in the lungs of a wide spectrum of mammalian species. They show how new molecular methodologies, and use of phylogenetics, have changed our views on the diversity and epidemiology of these fungal pathogens. *Pneumocystis* organisms have a marked host-species-related diversity, which is associated to close host specificity. High divergence among Pneumocystis species probably resulted from a prolonged process of coevolution with each mammal host and mostly associated with cospeciation like it was demonstrated in recent phylogenetic studies on primate-derived *Pneumocystis*. Therefore, *Pneumocystis* may be used to reconstruct the

phylogeny of mammal species, and possibly elucidate controversial issues in mammalian taxonomy. The third case by Kim Blasdell, Heikki Hentonnen, and Philippe Buchy relates on Hantavirus Genetic Diversity. Many hantaviruses are important zoonotic pathogens. The genus Hantavirus is a rapidly expanding group of viruses. Some members of the genus have been well known for over 30 years, but in the last 5 years at least 12 putative novel hantavirus species have been identified. Although the earliest hantaviruses to be recognized were usually identified in human, and rodent samples were often differentiated through serological techniques, in the past 20 years almost all of the hantavirus species were identified using molecular tools and their improvement has allowed a more precise characterisation of hantavirus taxonomy and species distribution.

The last part of the book – New integrations, includes drug resistance, which are part of the control strategy and emergence of new pathogens or new variants of older pathogens. The first chapter in this part by Raymond Ruimy, François Barbier, David Lebeaux, Etienne Ruppé and Antoine Andremont deals with nasal methicillin-resistant coagulase-negative Staphylococci: a reservoir of mecA gene for *Staphylococcus aureus*. The efficacy of the penicillin on the *Staphylococcus aureus* infection has justified its widespread use. As a consequence a rapid emergence of penicillin resistant strains has been recorded. The use of a new semi-synthetic β-lactam, the methicillin has allowed treating these resistant strains. However, as for the penicillin, meticillin resistant *S. aureus* (MRSA) have rapidly been selected but remained mainly confined to hospitals. Recently, MRSA have independently emerged in the community. The acquisition of mecA gene carried on a mobile genetic element called Staphylococcal Cassette Chromosom mec (SCCmec) confers meticillin resistance. This gene encodes for a penicillin-binding protein which has a low affinity not only for most semi-synthetic penicillin (such as methicillin) but also for the entire β-lactam class. SCCmec is widespread in coagulase negative staphylococcus (CoNS). Since meticillin susceptible *S. aureus* and CoNS share the same ecological niches in humans, transfer could occur from MR-CoNS to MSSA. The authors will argue in favour of this transfer hypothesis. A second chapter by Silvestre and Cabaret is related to nematode resistance to anthelmintics "Molecular knowledge of mechanisms of helminth resistance: Importance for diagnostic and epidemiology". Helminths infestations in animals and humans are mostly controlled by anthelmintics. Large-scale treatment has led to resistance to many drugs, namely benzimidazoles, avermectins and levamisole. Whereas avermectins and levamisole resistance appear to be multigenic, benzimidazole resistance is largely monogenic, based on polymorphism in β-tubulin gene. Spatial distribution of β-tubulin alleles in closed helminths populations supports a common origin of alleles on a specific farm. The main forces responsible for anthelmintic resistance development in field populations are the introduction of ancestral alleles (i.e. pre existing polymorphism in helminths before herds' constitution) and the selection of alleles appeared after herds' constitution. The third chapter, "Molecular epidemiology of disease resistance genes with perspectives for researches on biological invasions and hybrid zones" by Nathalie Charbonnel and Jean-François Cosson, considers the resistance of the host to infection. From the hosts' point of view, molecular epidemiology has lead to the identification of components of susceptibility and resistance to infectious

diseases. The authors aim to explain why resistance to infections exhibits such a remarkable degree of polymorphism while being resistant obviously confers a high selective advantage to hosts. In this context, they develop host molecular epidemiology with regard to the concepts of evolutionary biology and immunoecology. They detail the mechanisms that are likely to underlie the variable degrees of host resistance polymorphism observed among natural populations. Particular attention is given to recently emphasized topics, including the risks of immunopathology, the spatial structure of populations, the impact of neutral evolutionary processes and the phenotypic plasticity of resistance. They evidence the consequences of this polymorphism for disease epidemiology both from empirical examples and genetic epidemiological modelling of resistance evolution. They stress the numerous gaps that remain to be explored to understand these patterns of disease resistance polymorphism. A chapter by Gwenaël Vourc'h, Olivier Plantard, and Serge Morand questions on diversity and ecology of diseases "How does biodiversity influence the ecology of infectious disease?". Over the past years, biodiversity has been reduced on an unprecedented scale, while new infectious diseases are emerging at an increasing rate. Greater overall biodiversity could lead to a greater diversity of hosts and thus of pathogens. Yet disease regulation – due to the buffering role of host diversity – is considered to be one of the services provided by biodiversity. In this chapter, they ask how biodiversity is linked to infectious disease. First, they investigate the influence of the biodiversity of pathogens. They highlight that the number of pathogen species is not well known but that new findings are facilitated by the rapid expansion of molecular techniques. They emphasize that pathogen intraspecific diversity is a crucial factor in disease emergence and allows pathogens to adapt to the selective pressures they face. Second, they investigate how the diversity of hosts influences infectious disease ecology. For multi-host diseases, a change in host species richness or abundance can modify the dynamics of local infectious diseases by either reducing or increasing the risk of transmission to the targeted host species. Third, they rapidly examine the role of biodiversity in the treatment of infectious diseases. Alexandre Caron, Serge Morand and Michel de Garine-Wichatitsky investigate the interactions between hosts either wild, domesticated or humans in a perspective of global health: "Epidemiological interaction at the wildlife/livestock/human interface: Can we anticipate emerging infectious diseases in their hotspots? A framework for building emerging diseases processes in their hot spots". Emerging infectious diseases' hotspots have been identified as multi-host and multi-pathogen systems often characterized in tropical ecosystems by an extensive wildlife/domestic/human interface. The pathogen communities shared by the wild and domestic populations at this interface reflect the historical epidemiological interactions between them. The understanding of the mechanisms of pathogen transmission in a specific ecosystem can provide an interaction network between host populations defined by nodes and linkages and characterized by the frequency, intensity and direction of the interactions with a direct input for targeted disease surveillance.

This book encompasses most of the questions related to molecular epidemiology either in animals or humans. It does not pretend to an exhaustive coverage of molecular epidemiology, an expanding topic in science. We expect that we presented most of the important facets of this new research in epidemiology.

Chapter 2
Molecular Typing of Bacterial Pathogens: A Tool for the Epidemiological Study and Control of Infectious Diseases

Marie Hallin, Ariane Deplano, and Marc J. Struelens

Abstract Molecular typing is nowadays an integral part of the public health microbiology toolbox. It indexes subspecies genotypic or phenotypic characters to estimate the genetic relatedness of microbial isolates and infer from it their probability of belonging to the same chain of transmission. Typing is used both to investigate outbreaks and enhance the resolution of disease surveillance at different population levels: (i) locally, in hospitals or the community, by clinical or public health laboratories; (ii) nationally, by reference laboratories or (iii) globally, through international surveillance networks. This chapter provides an overview of currently available and emerging technologies for typing human bacterial pathogens, discusses their suitability to different levels of use and reviews examples of integrated typing in advanced surveillance systems.

2.1 Introduction

Microbial typing allows the differentiation of epidemiological related from unrelated isolates of the same bacterial species. It is used to elucidate the source and route of transmission of micro-organisms causing outbreaks of infectious diseases. It can also be applied to basic research including in depth investigations of infectious disease pathogenesis, bacterial population structures and microbial genetic diversity in diverse ecosystems (van Belkum et al. 2007).

M. Hallin (✉) • A. Deplano
Laboratoire de Microbiologie et Laboratoire de Référence des Staphylocoques-MRSA, Hôpital Erasme, 808 route de Lennik, 1070 Brussels, Belgium
e-mail: marie.hallin@ulb.ac.be

M.J. Struelens
Scientific Advice Unit, European Centre for Disease Prevention and Control, 17183 Stockholm, Sweden

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_2,

A typing method (TM) should ideally be rapid, robust and produce objective and portable data enabling subspecies discrimination of microbial isolates to trace pathogen transmission unambiguously across various population and time scales. Even though next generation whole genome sequencing has become the ultimate reference technology (Harris et al. 2010; Lewis et al. 2010), high cost and complex data analysis hurdles must still be overcome to make it routinely applicable for epidemiological typing. Therefore, less comprehensive microbial typing methods (TM) still have to be carefully chosen with respect to analytical performance and convenience criteria according to the pathogen and application of interest (Struelens and ESGEM 1996; van Belkum et al. 2007).

TM can be classified according to the nature of the character they explore, where "phenotypic" TM are opposed to "genotypic" TM or following their standardisation and portability of results they produce, where "comparative" TM are opposed to "definitive (library) techniques." A *comparative TM* assesses relatedness within a set of isolates without reference to previously characterised strains or type database. It allows comparison of isolates within a single experiment but the results are not easily comparable with previous or future samples or with other laboratories. An example of currently used comparative TM is pulsed-field gel electrophoresis (PFGE) typing according to in house protocol. In contrast, a *library TM*, generates type allocation of isolates according to standard nomenclature and allows exchange of results between laboratories and comparisons of results over time by integration into cumulative, shared databases. An example of library typing system is the spoligotyping scheme of *Mycobacterium tuberculosis*.

Phenotypic TM, such as biotyping, serotyping or antibiogram based typing, have been widely used in the past as library TM as they are relatively easy to implement, do not require costly equipment, have been well standardised and generate results that are simple to categorise and compare. Their major drawback is related to the characters they study, which are subject to regulation and variable expression according to in vitro or in vivo environmental conditions. Furthermore, phenotypic characters can undergo evolutionary convergence and consequently may not reflect accurately the genotype, nor provide a reliable epidemiological marker by falsely lumping together isolates of divergent descent**.**

Because phenotypic characters lack performance and reliability, *genotyping methods* based on comparison of restriction or amplification of DNA fragment patterns (DNA fingerprinting) or DNA sequence determination (DNA sequence based methods) are nowadays preferred. Sequence based methods include methods involving the sequencing of a single locus (SLST) or multiple loci (MLST).

2.2 Applications of Typing Methods

The choice of a TM in a given situation will depend on two main parameters: (1) the population structure of the bacterial species to type, and (2) the space and time scale of the population under study.

2.2.1 Population Genetics

Evolutionary speed, intra-species diversity and genetic population structure differ widely among human bacterial pathogens (Deplano and Struelens 2006). Depending on the relative frequency of recombination and point mutations, the population structure of a bacterial species can be localised across a continuum going from strict *clonality* (where daughter isolates are identical to the parents, issued from binary cell division, and diversification solely occurs by mutational genetic "drift") to *panmixis* (where daughter isolates vary extensively from the parents due to high rate of random intergenomic gene exchange). To date, the majority of bacterial pathogens studied display a predominantly clonal structure, characterized by the existence of dominant lineages, showing a strong "linkage disequilibrium" (non-random association of alleles at different loci) (Deplano and Struelens 2006; Musser et al. 1990; Smith et al. 1993; Tibayrenc 1996). However, in certain species with a panmictic population structure, transient clonality may also be found when lineages become ecologically isolated or when epidemic dissemination occurs faster than recombination within a population. For example, *Neisseria meningitidis*, which has typically a panmictic population structure, displays epidemic clones/clusters that can clearly be identified in geographically circumscribed areas (Deplano and Struelens 2006; Smith et al. 1993; Struelens and ESGEM 1996). On the opposite, independent evolution of unrelated lineages under the same selective pressure can lead to the acquisition of common phenotypic characters. This *convergence* phenomenon can blur epidemiological data based on TM targeting these characters and consequently lead to inaccurate clonality inference.

A good knowledge of the genetic population structure and molecular mechanisms of evolution of a given bacterial species is thus a prerequisite for selecting suitable TM and interpretation of results. For instance, the same TM can prove to be highly discriminating when applied to a panmictic species, but not discriminating enough when applied to a highly clonal species.

2.2.2 Study Scale

MT can be applied to diverse applications ranging from analysis of microbiota samples originating from an individual host up to samples from the world population. Within a single host, TM can be used to differentiate the recurrence of an infectious process from re-infection with a new isolate from the same species as cause of disease relapse. TM can also help to confirm the anatomic source of an endogenous infection such as intravenous cather-related bacteraemia.

The classical application of molecular typing is for supporting outbreak investigations by use of either comparative or library TM. In this context, TM can help to determine (i) the extent of epidemic spread of an infectious agent in an exposed population, (ii) the source or vector of epidemic contamination and (iii) the impact of control measures on the course of the epidemic and rate of strain transmission.

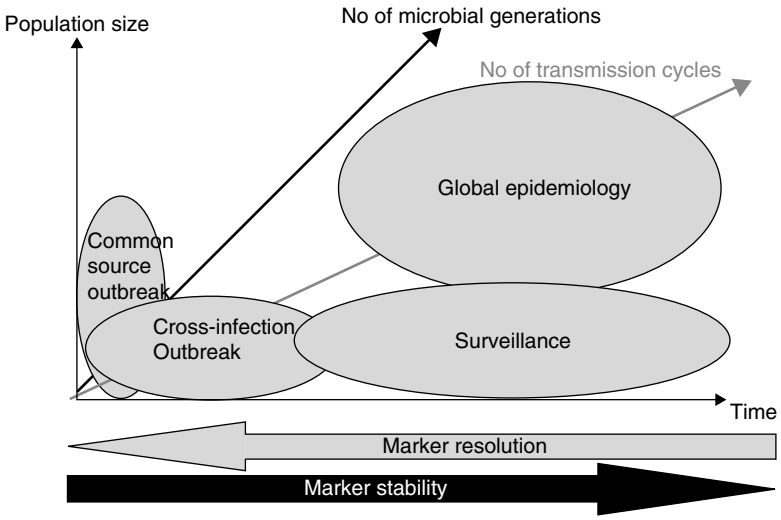

Fig. 2.1 Different time-space scales for tracking epidemic clones

Library TM are primarily used by reference laboratories for enhanced surveillance of infectious diseases, to provide information on pathogen type distribution frequency over time and place in relation to disease occurrence among populations.

Sequence based TM are also used for the study of genetic interrelationships in bacterial populations for evolutionary studies.

Finally, sequence based TM, comparative genomics in particular, are increasingly used to search for determinants of virulence or drug resistance, allowing identification of novel pathovar- or epidemic type diagnostic markers, or giving insight into microbial pathogenesis (van Belkum 2002).

These different application levels are important to distinguish since the "epidemiological scale" is a key parameter to consider when selecting the appropriate TM for a given purpose. For example, genotypic methods used to assess genetic structure in bacterial populations for evolutionary studies generally target moderately conserved loci like housekeeping genes. Therefore, they are very useful to calibrate the evolutionary clock speed of other TM but usually show insufficient discriminatory power in discriminating bacterial isolates collected over relatively short time periods from restricted geographical areas. High resolution methods are preferable for local outbreak studies. These TM target more rapidly evolving loci such as surface protein encoding genes or variable number short tandem repeat elements. Similarly, while comparative typing can be sufficient for outbreak investigation where the needs for reproducibility, standard nomenclature or stability over time are limited, national and international surveillance systems require library TM that allow comparison of typing data generated over years in different laboratories for monitoring the prevalence and spread of transmissible clones. For local epidemiological surveillance

purpose, a compromise must be found between discriminatory power and marker stability for selecting the appropriate TM (Fig. 2.1).

2.3 Performance Criteria of Typing Methods

Prior to use for epidemiological purpose, a microbial typing system should be evaluated and validated. Detailed recommendations for optimal use and quality assessment of typing methods have been published by the ESCMID Study Group on Epidemiological Markers (ESGEM) (van Belkum et al. 2007). They propose a number of performance criteria, including:

- *Typeability*. This is the ability of the TM to assign a type to isolates tested, expressed as the percentage of typeable isolates over the total number of typed isolates.
- *Stability*. This criterion refers to the stability of the marker used for typing within the species typed. Usually, this is tested *in vitro*, by typing serial subcultures of several strains. Ideally, stability should also be assessed in vivo. Since recombination and mutation rates differ by species, the stability should be evaluated for each species studied.
- *Reproducibility*. This is the ability of the TM to assign the same type to the same isolate on repeat testing. Reproducibility may be influenced by many parameters involved in the typing protocol, including interpretation of results. All these potential biases have to be assessed during evaluation. Additionally, reference type controls should be included to check each run quality. On a larger dimension, *inter-laboratory reproducibility* is a key to enable the exchange of data through laboratory networks. This implies the adoption of consensus protocols and regular participation of network members to external proficiency testing schemes.
- *Discriminatory power*. This reflects the capacity of the TM to assign two different types to two epidemiologically unrelated strains in the population studied. This capacity can be expressed using the Simpson index of diversity (DI) (Hunter and Gaston 1988), which expresses the probability to assign two different types to two strains sampled randomly from a large collection of epidemiologically unrelated isolates representative of the ecological niche studied (Struelens and ESGEM 1996). The determination of a confidence interval enables the comparison of DI of several TM (48).
- *Epidemiological concordance*. This criterion refers to the ability of the TM to classify in identical or related types “single clone derived” isolates. This should be tested on several well documented sets of epidemiologically-related isolates from independent outbreaks.

Finally, several convenience criteria including the flexibility/versatility (applicability to a maximum range of pathogen species with minimal protocol modification), rapidity, ease of use, technical accessibility, equipment and reagent costs, and last but not least, interpretation easiness and exportability of results must be taken into account when choosing a TM.

2.4 Overview of Currently Available Typing Methods

2.4.1 Phenotypic Typing Methods

2.4.1.1 Antibiogram-Based Typing

Antimicrobial susceptibility testing is performed by clinical microbiology laboratories to guide chemotherapy. This method can be used for typing if appropriate drugs are selected and if resistance patterns are sufficiently discriminating and stable for use as a clonal marker. Indeed, some resistance determinants are located on plasmids that can be lost at high frequency. Moreover, resistance expression can be also controlled by complex regulatory systems. Multivariate analysis of susceptibility patterns based on diameter inhibition zone has been successfully used for MRSA typing and showed good correlation with ribotyping (Blanc et al. 1996). International databases including antibiograms and data on clinical and geographical origins of isolates were used to estimate the incidence and spread of specific resistance markers or multiresistant clones(Fluit et al. 2006).

2.4.1.2 Serotyping

Serotyping systems are widely used as a reference method for typing *Salmonella* and *E. coli* isolates. Typeability and discrimination are variable and interpretation is often complicated by antigen cross-reaction phenomena. This library TM is used in reference laboratories with standardised reagents and method. Discrimination of serotyping depends on the species. Instable serotypes occur secondary to recombination events (for examples: transformation in *S. pneumoniae*, transduction in *Salmonella*). Overall, congruence of antigenic variants with the population genetic structure is poor, even in species with clonal structure like *E. coli* (Ochman and Selander 1984) or *Legionella pneumophila* (Cazalet et al. 2008). Some serotyping schemes are now being replaced by their genotypic equivalents. For example, restriction analysis of the amplified O-antigen gene cluster has proven an interesting alternative to classic serotyping of *E. coli* (Coimbra et al. 2000).

2.4.1.3 Phage Typing

It determines the lytic pattern of bacterial isolates exposed to a defined set of phages. This method was developed for a small number of pathogenic species, although time-consuming and technically demanding procedure. This and the necessity for keeping stocks of phages and propagating strains, confined it to reference laboratories. It shows variable discrimination and is mostly limited by poor typeability and reproducibility. Due to these limitations, phage typing has been replaced by molecular TM. Acquisition or loss of phages, which play a role in virulence, can be traced by DNA analysis, providing a modern extension of phage typing (Goerke et al. 2004).

2.4.1.4 Multilocus Enzyme Electrophoresis (MLEE)

This phenotype based genotyping method identifies electrophoretic variants of a set of 10–35 housekeeping enzymes. MLEE has once been considered as the reference method for defining the phylogenetic structure of lineages in clonal bacterial populations (Musser et al. 1990). MLST, the "genotypic equivalent" of MLEE, has replaced it.

2.4.1.5 Mass Spectrometry

Mass spectrometry (MS) was initially used for analysing the composition of complex mixtures of chemical or organic (macro) molecules. *The Matrix-assisted laser desorbtion ionization time-of-flight* (MALDI-TOF), can be applied to whole cells, allowing the characterisation of their constituent molecules as complex spectra that can be compared to mass spectra databases. MALDI-TOF has been recently validated as a fast and reliable tool for bacterial species identification (Cherkaoui et al. 2010; Stevenson et al. 2010). The spectra generated can also present sufficient differences within species to allow to differentiate strain types, such as distinct isolates of *E. coli* O157:H7 (Everley et al. 2008), or *Legionella pneumophila* (Fujinami et al. 2010). Another promising application of MALDI-TOF MS technology is its capability of SNPs detection in PCR products (Tang et al. 2004).

Another spectrometric method, called *Raman spectroscopy*, which analyses the reflected light of intact cells, has been commercially developed for the identification of microorganisms at the species as well as at subspecies level (Maquelin et al. 2006; Willemse-Erix et al. 2010).

2.4.2 Genotypic Typing Methods

The performance profile of genotypic methods used for typing of bacterial pathogens is compared in Table 2.1.

2.4.2.1 Fragment Based TM

Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphism (RFLP) of chromosomal DNA using frequently cutting enzymes coupled with hybridisation with labelled probes was the first generation bacterial genotyping method used in the 1980s. A successful use of this approach is illustrated by IS*6110* typing of *M. tuberculosis* which was standardized over 15 years ago and since used worldwide with a central integrated database. During the past two decades, pulsed field gel electrophoresis (PFGE) protocols of macrorestriction DNA profiles have been developed and established as the "gold

Table 2.1 Comparison of performance profile of molecular epidemiological typing systems

Methods	Discriminatory power	Reproducibility	Polyvalence	Objectivity of data	Portability of data	Throughput
Fingerprint-based:						
RFLP+probe	Good	Good	Good	Good	Poor	Low
PFGE	Excellent	Good	Excellent	Poor to good	Poor	Low
MLVA	Good	Good	Excellent	Good	Poor	Medium
Rep-PCR	Good	Poor to good	Excellent	Poor	Poor	Medium
Sequence-based:						
SLST	Poor-good	Excellent	Poor	Excellent	Excellent	Medium
MLST	Good	Excellent	Excellent	Excellent	Excellent	Low
SNP	Excellent	Excellent	Excellent	Excellent	Excellent	High
Whole genome	Optimal	Excellent	Excellent	Excellent	Excellent	Low
Hybridization:						
Oligotyping	Good	Excellent	Poor	Good	Good	Medium
Microarray	Excellent	Good	Poor	Good	Good	High

standard" TM for many food-borne and nosocomial bacterial pathogens. PFGE takes advantage of using infrequently cutting (or "macro") restriction enzymes that cleaves whole bacterial genomes in 10–30 DNA fragments. These large DNA fragments are separated according to size by periodically alternating the direction ("pulsing") of the electric field in specially designed electrophoresis systems (PFGE). Some major strengths of PFGE include its versatility and high discriminatory power that are well documented. However, this technology suffers from some drawbacks: it is time- consuming and laborious to perform, requires a high quality, high concentration DNA preparation and rigorous standardization of electrophoresis protocols to enable inter-laboratory comparison (Murchan et al. 2003; van Belkum et al. 1998). Centralized PFGE databases were developed for "real time" typing of food-borne pathogens by the PulseNet public health laboratory programme in the USA and later extended worldwide (Swaminathan et al. 2001). This PFGE standardisation process has been more limited for hospital-associated pathogens such as methicillin-resistant *S. aureus* (MRSA) because of lack of reproducibility among laboratories (van Belkum et al. 1998). Despite this difficulty, reference PFGE typing protocols, databases and nomenclature of MRSA epidemic types were developed in the U.S.A and Europe (Hallin et al. 2007; McDougal et al. 2003; Murchan et al. 2003).

PCR Fingerprinting

Repetitive DNA sequences are interspersed throughout the genome of all bacteria. Amplification of short genomic fragments lying between repetitive elements by using low-stringency PCR with outward primers, called rep-PCR, is an efficient method for bacterial typing (Deplano et al. 2000; van Belkum et al. 1996). Its major advantages include flexibility, technical simplicity and rapidity. However, rep-PCR patterns are difficult to interpret and non-exchangeable between laboratories due to variable fragment amplification efficiency (Deplano et al. 2000). This method was therefore considered a comparative system applicable as in house "clonal screen" method. Recently, a semi-automated rep-PCR based system (Diversilab™, Bacterial Barcodes, USA; distributed by BioMérieux, France) was successfully evaluated for identification of the *E. coli* O25:H4-ST131 clonal lineage (Bonacorsi et al. 2009; Pitout et al. 2009) and typing of *Acinetobacter baumanii* isolates (Carretto et al. 2008; Fontana et al. 2008). It was found to be less discriminating than PFGE for *S. aureus* typing (Tenover et al. 2009). This system is well standardized, technically simple and reproducible but relatively expensive.

Multilocus Variable Number of Tandem Repeat Analysis (MLVA)

Repetitive DNA is widely found in prokaryotes (van Belkum et al. 1998). Repeats elements vary in size, location and complexity. Elements which are organized in tandem, are located at a single locus in the genome and show variation in repeat unit

number, are designated variable number of tandem repeat (VNTR) loci. Multiple locus VNTR analysis (MLVA) involves the determination of the number of repeat copy units in multiple loci by using multiplex PCR for size analysis of repeat regions. Size determination of PCR products is performed by agarose electrophoresis or higher resolution systems such as capillary electrophoresis and mass spectrometry. For each locus, each allele type representing the number of repeats is assigned a digit code. Allele types from multiple loci are combined to designate a multi-digit MLVA type. This strategy is of particular interest for high resolution genotyping of pathogens with a high degree of genome homogeneity, such as *Bacillus anthracis* and *M. tuberculosis*. MLVA systems have been recently developed for several human pathogens, including potential agents of bioterrorism (van Belkum et al. 2007; van Belkum 2007). In general, MLVA shows good correlation with other genotyping methods and its technical simplicity and rapidity make it an attractive TM. A potential disadvantage of MLVA typing is the pace of evolution of repetitive DNA that may be too rapid thereby compromising epidemiological concordance. Thus, the variability of each VNTR must be assessed carefully. An example of this validation is given by VNTR analysis of different classes of interspaced genetic elements, called mycobacterial interspersed repetitive units (MIRUs), for epidemiological typing of *M. tuberculosis* (Supply et al. 2001, 2006). Among numerous sets of MIRU-VNTR loci which have been described, an optimised 15-locus system was validated as a new standard for routine epidemiological discrimination of *M. tuberculosis* isolates and a 24-locus system was further developed as higher resolution tool for phylogenetic studies (Oelemann et al. 2007; Supply et al. 2006). The method proved superior to previous standard methods such as *IS*6110 in terms of typeability and to spoligotyping in terms of discriminatory power (Allix et al. 2006).

2.4.2.2 Sequence Based Typing Methods

Single Locus Sequencing Typing (SLST)

One key advantage of all sequence-based methods over "band-based" techniques is that sequence data is non ambiguous and portable between laboratories. Single locus sequence typing (SLST) analyses a single variable genetic locus to save workload, cost and time. The SLST approach is well validated for typing of *S. aureus* based on analysis of an hypervariable region of the staphylococcal protein A gene (*spa* gene). This polymorphic X region contains a variable number of 24 bp repeat regions. Typing based on sequence comparison of this X region (*spa* typing) combines practical advantages, such as rapidity, reproducibility, ease of use and portability (Aires-de-Sousa et al. 2006; Koreen et al. 2004). The X region indexes micro- and macro-variations that appear highly congruent with whole genome clonal variation. This enables use of *spa* typing both for local and global epidemiology (Koreen et al. 2004). Recent evaluation of potential use of *spa* typing for long-term nationwide epidemiology of *S. aureus* showed high discriminatory power and

achieved high concordance with PFGE and MLST analysis. The development of a dedicated software with the possibility of rapid assignation of sequence led to an automated system showing 100% inter-center reproducibility (Aires-de-Sousa et al. 2006; Harmsen et al. 2003). *spa* typing has been successfully evaluated as automated clonal alert system to detect local MRSA outbreak (Mellmann et al. 2006). Occasional "violation" of MLST clonal complex assignment by *spa* typing was nevertheless observed, suggesting that *spa* typing should preferably be used in combination with additional markers, such as staphylococcal cassette chromosome *mec* type to ensure reliable clonal typing (Hallin et al. 2007). Another epidemiologically relevant and conceptually similar SLST scheme is *emm* sequence typing of *S. pyogenes* (Facklam et al. 1999). This TM analyses variation in the surface-associated streptococcal virulence protein M. An international database including updated *emm* types correlated with corresponding M-serotypes is also publicly available on internet (www.cdc.gov/ncidod/biotech/strep/strepindex.htm).

Multi Locus Sequencing Typing (MLST)

MLST indexes variation in loci encoding several housekeeping genes. It was originally designed to identify hyper-virulent lineages of *Neisseria meningitidis* (Maiden et al. 1998). MLST detects variation in the sequence of DNA fragments (400–500 bp in size) of five to ten housekeeping genes. Each gene sequence is considered as an allele type. The allele type for each of the gene sequence is combined into an allelic profile defining the sequence type (ST). MLST has been applied to a variety of human bacterial pathogens including *S. aureus, S. pneumoniae* and enterococci, and forms the basis of current nomenclature to label hyper-virulent and antibiotic-resistant clones of these organisms. MLST databases are centrally curated and made publicly available over the Internet to ensure that new types follow uniform nomenclature. MLST is a universal and definitive TM that is now recognized as the "gold standard" for defining the subspecies genetic population structure of many microorganisms. Currently, MLST schemes are available for more than 35 pathogenic bacteria and fungi (www.mlst.net, www.pasteur.fr/recherche/genopole/PF8/mlst, mlst.ucc.ie, pubmlst.org).

MLST data have been employed in epidemiological investigations of various scales and in studies of the population biology, pathogenicity, and evolution of bacteria. It is important to remember that this method was developed to assess genetic interrelationships in bacterial populations for evolutionary studies. These studies identified bacterial populations with house keeping genes showing high or moderate diversity. In contrast, MLST is less discriminating for epidemiological investigations such as those of hospital outbreaks that include bacterial isolates collected over a short time period from a restricted geographical area. Because it implies both strand sequencing of multiple genes, MLST is a labour and time consuming and expensive TM, thereby restricting its routine use for high throughput epidemiological typing. In case of national or international studies, MLST is useful for clonal type identification according to standard nomenclature.

For some pathogens like *Legionella pneumophila*, specially designed multi-variable locus sequence typing schemes have been developed for high resolution epidemic investigations. They have the advantages over other fingerprinting methods of being fully portable and directly applicable to clinical samples without the need for culture of this fastidious organism (Ginevra et al. 2009).

Single Nucleotide Polymorphism (SNP) Genotyping

Single nucleotide polymorphism (SNP) is a change in a single nucleotide in one sequence relative to wild type due to random nucleotide mutation, horizontal gene transfer or intragenic recombination. SNP genotyping methods are primarily applied to define relationship among isolates of highly homogeneous pathogens such as *M. tuberculosis*, *E. coli* O157:H7 or *B. anthracis*. Detection of the most frequently occurring SNPs associated with ST allelic variants is a simple and cost-effective alternative to full MLST. SNP systems offer similar phylogenetic information as MLST schemes on which they are based but they show less discrimination and are incapable to detect new allelic types. SNP genotyping can be applied to screen specific epidemic or virulent clones suitable for surveillance. For instance, conserved SNPs which characterize the allelic profile of the major epidemiological lineage ST-21 of *Campylobacter jejuni* were identified (Best et al. 2004). Clonal complex ST-21 is one of the largest CC constituting 26% of all submitted isolates (http://pubmlst.org/campylobacter). It is frequently associated with human disease and food chain sources (Dingle et al. 2001).

2.4.2.3 Hybridisation Methods

Direct Hybridization

In these methods, the microbial DNA to be investigated is immobilized on a support and probed with specific DNA molecules. Multiple hybridization technologies do exist. Hybridization can be made directly on genomic DNA such as in the binary typing scheme developed with PCR-generated, clone-specific probes for epidemiological typing of *S. aureus* (van Leeuwen et al. 2001).

One of the earliest broad-range bacterial genotyping system invented was ribotyping. Here, probes targeting 16S and 22S rDNA are used to estimate the number and position of ribosomal loci in the bacterial genome. In this case, hybridization is made after specific enzymatic restriction of chromosomal DNA and transfer on a solid support. This method is highly reproducible and has a moderate discriminatory power. Fully automated RiboPrinter™ (Qualicon, Inc., USA) allows reduction of hand-on time and comparison of profiles among laboratories. It was adopted for some pathogens important in food microbiology even tough it is more expensive and less discriminating than other genotypic methods (Pfaller et al. 1996).

Hybridization can be also made on amplified DNA products such as for *M. tuberculosis* spoligotyping. In this case, the PCR amplified target DNA is a

locus harbouring tandem repeats with some internal sequence variation. Specific variants are typed by dot-blot hybridization using an array of repeat-specific oligonucleotide DNA probes. The major advantage of spoligotyping over IS*6110* typing is that can be applied directly to clinical samples, without the need of prior culture (Goyal et al. 1997; Kamerbeek et al. 1997).

DNA Microarray Typing

Since microbiology has entered the postgenomic era, bacterial typing is increasingly based on analysis of complete genomes. Although whole genome sequencing remains costly and limited to laboratories with advanced DNA sequencing platforms, DNA microarrays are now available and provide extensive, if not comprehensive, information on whole gene content for bacterial genotyping. Microarrays are collections of microscopic probes (usually DNA probes) which can be hybridized with labelled target molecules (DNA or RNA) on a solid substrate such as glass slide to produce quantitative or qualitative heteroduplex signals. Various types of microarrays are classified upon their technical characteristics such as the nature of the probe, the type of solid support or the method used for target detection (Miller and Tang 2009). Microarrays can be used for a wide range of applications including typing. The major advantage is the ability of this technique to extensively characterize and simultaneously genotype bacterial isolates. Examples of bacterial pathogens for which microarray has been successfully used for epidemiological typing include *Salmonella enterica* (Majtan et al. 2007), *Staphylococcus aureus* (Monecke et al. 2007), pathogenic *Escherichia coli* and *Campylobacter jejuni* (Garaizar et al. 2006). As the technology improves, the use of microarray platforms is moving from research to clinical and public health laboratory applications.

High-Throughput Genomics Typing

The proof of principle of applying second generation DNA sequencing technologies to epidemiological investigation of infectious diseases has been provided with a recent analysis of the global epidemiology and local microevolution of a successful clone of MRSA (Harris et al. 2010). These comparative genomics data indicated a core gene mutation rate of one SNP every 6 weeks during patient-to-patient transmission within a hospital ward. They also confirmed the occasional occurrence of intercontinental clone spread followed by regional expansion of subclonal variants. It is likely that high-throughput sequencing of whole bacterial genomes will become increasingly used to decipher the full extent of population genomics of clinically relevant bacterial species at local, regional and global levels. Once the molecular clocks of all informative SNPs will be better ascertained within lineages, interpretation of comparative genomics will be possible for resolving outbreak investigations and mapping pathogen transmission.

2.5 Molecular Typing as a Public Health Tool

The increasing access to routine molecular typing at clinical or public health laboratory level has enhanced the sensitivity of epidemiological surveillance for early detection and warning of epidemic diseases. For instance, rapid detection of multi-resistant bacteria belonging to previously described hypervirulent and highly transmissible clonal lineages can inform infection local control measures such as patient isolation and contact screening among critically-ill or immuno-compromised patients (Deplano et al. 2007a, b).

Molecular typing is now forming an integral part of the microbiology services offered by public health reference laboratories. By integrating pathogen genotypic data with epidemiological data of clinical cases, it enhances detection of previously unrecognised clusters of disease that may disseminate on a global scale, such as travel-acquired legionnaires's disease or food-borne infections (Musser et al. 1990; Miller and Tang 2009). An important prerequisite for reliable use of any molecular typing system for guiding infection control strategies is the organisation of external quality assessment trials and inter-centre validation studies for typing network participants (Aires-de-Sousa et al. 2006; Deplano et al. 2006; van Leeuwen et al. 2002). This, however, represents a substantial challenge as new typing technologies appear and improve at an accelerating rate. Whereas it is clear that sequence-based technologies offer substantial superiority to other currently available typing approaches, it remains to be seen whether high throughput genome sequencing will either supersede or help design more parsimonious yet micro-evolutionary informative SNP detection strategies.

References

Aires-de-Sousa M, Boye Z, de Lencastre H et al (2006) High interlaboratory reproducibility of DNA sequence-based typing of bacteria in a multicenter study. J Clin Microbiol 44:619–621

Allix C, Walravens K, Saegerman C et al (2006) Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of *Mycobacterium bovis* and comparison of the method with IS6110 restriction fragment length polymorphism analysis and spoligotyping. J Clin Microbiol 44:1951–1962

Best EL, Fox AJ, Frost JA et al (2004) Identification of *Campylobacter jejuni* multilocus sequence type ST-21 clonal complex by single-nucleotide polymorphism analysis. J Clin Microbiol 42: 2836–2839

Blanc DS, Petignat C, Moreillon P et al (1996) Quantitative antibiogram as a typing method for the prospective epidemiological surveillance and control of MRSA: comparison with molecular typing. Infect Control Hosp Epidemiol 17:654–659

Bonacorsi S, Bidet P, Mahjoub F et al (2009) Semi-automated rep-PCR for rapid differentiation of major clonal groups of *Escherichia coli* meningitis strains. Int J Med Microbiol 299:402–409

Carretto E, Barbarini D, Farina C et al (2008) Use of the DiversiLab semiautomated repetitive-sequence-based polymerase chain reaction for epidemiologic analysis on *Acinetobacter baumannii* isolates in different Italian hospitals. Diagn Microbiol Infect Dis 60:1–7

Cazalet C, Jarraud S, Ghavi-Helm Y et al (2008) Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. Genome Res 18:431–441

Cherkaoui A, Hibbs J, Emonet S et al (2010) Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level. J Clin Microbiol 48:1169–1175

Coimbra RS, Grimont F, Lenormand P et al (2000) Identification of *Escherichia coli* O-serogroups by restriction of the amplified O-antigen gene cluster (rfb-RFLP). Res Microbiol 151:639–654

Deplano A, Struelens M (2006) Marqueurs épidemiologiques. In: Freney J, Renaud F, Bollet C, Leclercq R (eds) Actualités permanentes en Bactériologie Clinique. ESKA, Paris

Deplano A, Schuermans A, Van Eldere J et al (2000) Multicenter evaluation of epidemiological typing of methicillin-resistant *Staphylococcus aureus* strains by repetitive-element PCR analysis. The European study group on epidemiological markers of the ESCMID. J Clin Microbiol 38:3527–3533

Deplano A, de Mendonca R, de Ryck R et al (2006) External quality assessment of molecular typing of *Staphylococcus aureus* isolates by a network of laboratories. J Clin Microbiol 44:3236–3244

Deplano A, Denis O, Nonhoff C et al (2007a) Outbreak of hospital-adapted clonal complex-17 vancomycin-resistant *Enterococcus faecium* strain in a haematology unit: role of rapid typing for early control. J Antimicrob Chemother 60:849–854

Deplano A, Rodriguez-Villalobos H, Glupczynski Y et al (2007b) Emergence and dissemination of multidrug resistant clones of *Pseudomonas aeruginosa* producing VIM-2 metallo-beta-lactamase in Belgium. Euro Surveill 12:E070118

Dingle KE, Colles FM, Wareing DR et al (2001) Multilocus sequence typing system for *Campylobacter jejuni*. J Clin Microbiol 39:14–23

Everley RA, Mott TM, Wyatt SA et al (2008) Liquid chromatography/mass spectrometry characterization of *Escherichia coli* and *Shigella* species. J Am Soc Mass Spectrom 19:1621–1628

Facklam R, Beall B, Efstratiou A et al (1999) *emm* typing and validation of provisional M types for group A streptococci. Emerg Infect Dis 5:247–253

Fluit AC, van der Bruggen JT, Aarestrup FM et al (2006) Priorities for antibiotic resistance surveillance in Europe. Clin Microbiol Infect 12:410–417

Fontana C, Favaro M, Minelli S et al (2008) *Acinetobacter baumannii* in intensive care unit: a novel system to study clonal relationship among the isolates. BMC Infect Dis 8:79

Fujinami Y, Kikkawa HS, Kurosaki Y et al (2010) Rapid discrimination of *Legionella* by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Microbiol Res 166:77–86

Garaizar J, Rementeria A, Porwollik S (2006) DNA microarray technology: a new tool for the epidemiological typing of bacterial pathogens? FEMS Immunol Med Microbiol 47:178–189

Ginevra C, Lopez M, Forey F et al (2009) Evaluation of a nested-PCR-derived sequence-based typing method applied directly to respiratory samples from patients with Legionnaires' disease. J Clin Microbiol 47:981–987

Goerke C, Papenberg S, Dasbach S et al (2004) Increased frequency of genomic alterations in *Staphylococcus aureus* during chronic infection is in part due to phage mobilization. J Infect Dis 189:724–734

Goyal M, Saunders NA, Van Embden JDA et al (1997) Differentiation of *Mycobacterium tuberculosis* isolates by spoligotyping and IS6110 restriction fragment length polymorphism. J Clin Microbiol 35:647–651

Hallin M, Deplano A, Denis O et al (2007) Validation of pulsed-field gel electrophoresis and spa typing for long-term, nationwide epidemiological surveillance studies of *Staphylococcus aureus* infections. J Clin Microbiol 45:127–133

Harmsen D, Claus H, Witte W et al (2003) Typing of methicillin-resistant *Staphylococcus aureus* in a university hospital setting by using novel software for spa repeat determination and database management. J Clin Microbiol 41:5442–5448

Harris SR, Feil EJ, Holden MT et al (2010) Evolution of MRSA during hospital transmission and intercontinental spread. Science 327:469–474

Hunter PR, Gaston MA (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26:2465–2466

Kamerbeek J, Schouls L, Kolk A et al (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35:907–914

Koreen L, Ramaswamy SV, Graviss EA et al (2004) spa typing method for discriminating among *Staphylococcus aureus* isolates: implications for use of a single marker to detect genetic micro- and macrovariation. J Clin Microbiol 42:792–799

Lewis T, Loman NJ, Bingle L et al (2010) High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. J Hosp Infect 75:37–41

Maiden MC, Bygraves JA, Feil E et al (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 95:3140–3145

Majtan T, Majtanova L, Timko J et al (2007) Oligonucleotide microarray for molecular characterization and genotyping of *Salmonella spp.* strains. J Antimicrob Chemother 60:937–94

Maquelin K, Dijkshoorn L, van der Reijden TJ et al (2006) Rapid epidemiological analysis of *Acinetobacter* strains by Raman spectroscopy. J Microbiol Methods 64:126–131

McDougal LK, Steward CD, Killgore GE et al (2003) Pulsed-field gel electrophoresis typing of oxacillin-resistant *Staphylococcus aureus* isolates from the United States: establishing a national database. J Clin Microbiol 41:5113–5120

Mellmann A, Friedrich AW, Rosenkotter N et al (2006) Automated DNA sequence-based early warning system for the detection of methicillin-resistant *Staphylococcus aureus* outbreaks. PLoS Med 3:e33

Miller MB, Tang YW (2009) Basic concepts of microarrays and potential applications in clinical microbiology. Clin Microbiol Rev 22:611–633

Monecke S, Kuhnert P, Hotzel H et al (2007) Microarray based study on virulence-associated genes and resistance determinants of *Staphylococcus aureus* isolates from cattle. Vet Microbiol 125:128–140

Murchan S, Kaufmann ME, Deplano A et al (2003) Harmonization of pulsed-field gel electrophoresis protocols for epidemiological typing of strains of methicillin-resistant *Staphylococcus aureus*: a single approach developed by consensus in 10 European laboratories and its application for tracing the spread of related strains. J Clin Microbiol 41:1574–1585

Musser JM, Kroll JS, Granoff DM et al (1990) Global genetic structure and molecular epidemiology of encapsulated *Haemophilus influenzae*. Rev Infect Dis 12:75–111

Ochman H, Selander RK (1984) Evidence for clonal population structure in *Escherichia coli*. Proc Natl Acad Sci USA 81:198–201

Oelemann MC, Diel R, Vatin V et al (2007) Assessment of an optimized mycobacterial interspersed repetitive- unit-variable-number tandem-repeat typing system combined with spoligotyping for population-based molecular epidemiology studies of tuberculosis. J Clin Microbiol 45:691–697

Pfaller MA, Wendt C, Hollis RJ et al (1996) Comparative evaluation of an automated ribotyping system versus pulsed-field gel electrophoresis for epidemiological typing of clinical isolates of *Escherichia coli* and *Pseudomonas aeruginosa* from patients with recurrent gram-negative bacteremia. Diagn Microbiol Infect Dis 25:1–8

Pitout JD, Campbell L, Church DL et al (2009) Using a commercial DiversiLab semiautomated repetitive sequence-based PCR typing technique for identification of Escherichia coli clone ST131 producing CTX-M-15. J Clin Microbiol 47:1212–1215

Smith JM, Smith NH, O'Rourke M et al (1993) How clonal are bacteria? Proc Natl Acad Sci USA 90:4384–4388

Stevenson LG, Drake SK, Murray PR (2010) Rapid identification of bacteria in positive blood culture broths by matrix-assisted laser desorption ionization-time of flight mass spectrometry. J Clin Microbiol 48:444–447

Struelens MJ, ESGEM (1996) Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect 2:2–11

Supply P, Lesjean S, Savine E et al (2001) Automated high-throughput genotyping for study of global epidemiology of *Mycobacterium tuberculosis* based on mycobacterial interspersed repetitive units. J Clin Microbiol 39:3563–3571

Supply P, Allix C, Lesjean S et al (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. J Clin Microbiol 44:4498–4510

Swaminathan B, Barrett TJ, Hunter SB et al (2001) PulseNet: the molecular subtyping network for foodborne bacterial disease surveillance, United States. Emerg Infect Dis 7:382–389

Tang K, Oeth P, Kammerer S et al (2004) Mining disease susceptibility genes through SNP analyses and expression profiling using MALDI-TOF mass spectrometry. J Proteome Res 3:218–227

Tenover FC, Gay EA, Frye S et al (2009) Comparison of typing results obtained for methicillin-resistant *Staphylococcus aureus* isolates with the DiversiLab system and pulsed-field gel electrophoresis. J Clin Microbiol 47:2452–2457

Tibayrenc M (1996) Towards a unified evolutionary genetics of microorganisms. Annu Rev Microbiol 50:401–429

van Belkum A (2002) Molecular typing of micro-organisms: at the centre of diagnostics, genomics and pathogenesis of infectious diseases? J Med Microbiol 51:7–10

van Belkum A (2007) Tracing isolates of bacterial species by multilocus variable number of tandem repeat analysis (MLVA). FEMS Immunol Med Microbiol 49:22–27

van Belkum A, Sluijuter M, de Groot R et al (1996) Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. J Clin Microbiol 34:1176–1179

van Belkum A, van Leeuwen W, Kaufmann ME et al (1998) Assessment of resolution and intercenter reproducibility of results of genotyping *Staphylococcus aureus* by pulsed-field gel electrophoresis of SmaI macrorestriction fragments: a multicenter study. J Clin Microbiol 36:1653–1659

van Belkum A, Tassios PT, Dijkshoorn L et al (2007) Guidelines for the validation and application of typing methods for use in bacterial epidemiology. Clin Microbiol Infect 13(Suppl 3):1–46

van Leeuwen W, Libregts C, Schalk M et al (2001) Binary typing of *Staphylococcus aureus* strains through reversed hybridization using digoxigenin-universal linkage system-labeled bacterial genomic DNA. J Clin Microbiol 39:328–331

van Leeuwen WB, Snoeijers S, Werken-Libregts C et al (2002) Intercenter reproducibility of binary typing for *Staphylococcus aureus*. J Microbiol Methods 51:19–28

Willemse-Erix HF, Jachtenberg J, Barutci H et al (2010) Proof of principle for successful characterization of methicillin-resistant coagulase-negative staphylococci isolated from skin by use of Raman spectroscopy and pulsed-field gel electrophoresis. J Clin Microbiol 48:736–740

Part II
Contribution of Population Genetics

Chapter 3
Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases

Michel Tibayrenc

Abstract This chapter does not attempt to provide a comprehensive catalogue of the various methods of analysis and technologies used to characterize pathogens (molecular epidemiology). Rather it argues why molecular epidemiology should not be conceived without an extensive use of the concepts of population genetics and evolutionary biology. Moreover, it stresses that characterizing pathogens should open up to evaluating the impact of pathogens' genetic diversity on their relevant medical properties (downstream studies). Lastly, it presents the foreseeable future developments in this field, which has been upset by the exponential development of megatechnologies (massive sequencing, postgenomic studies, and bioinformatics).

3.1 Introduction

Before attempting to precisely define molecular epidemiology, it is interesting to observe it as a today's hot item. Using the term for a search in the SCOPUS database produces more than 15,000 references, most of them cited in the last few years. The term chiefly refers to transmissible diseases, but not exclusively. Some references concern other diseases such as cancer.

Considering the rapid growth of this field, there is no doubt that the publication of this book is quite timely. Actually, the premises of this burst onto the front stage were already apparent years ago. The field is at the crossroads of two major lines in modern biomedical research: the exponential development of molecular technologies and the preoccupying wave of emerging and reemerging infectious

M. Tibayrenc (✉)
Maladies Infectieuses et Vecteurs Ecologie, Génétique, Evolution et Contrôle,
MIVEGEC/IDVEGEC (UM1-CNRS 5290-IRD 224), IRD Center,
BP 64501, 34374 Montpellier Cedex 5, France
e-mail: Michel.Tibayrenc@ird.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_3,

diseases ("The golden age of genetics and the dark age of infectious diseases" Tibayrenc 2001). As for the threat of infectious diseases, it is worth noting that the alarms sounded these last few years (SARS, avian flu) have not, in the end, been confirmed as major pandemics. Still, the fact remains that transmissible diseases remain by far the major killers of the human species on a worldwide basis. It can also be said that they more than ever constitute the major selection pressure of humankind (Haldane 1949), since many of them kill at the age of reproduction, or before it. Even when considering industrial countries, much more than 40 years ago, today infectious diseases are a major concern with the problems of AIDS, nosocomial infections, antibiotic resistance and multidrug-resistant pathogens. Obviously, considering that transmissible diseases are under control would be quite unrealistic. Can it be considered therefore that molecular epidemiology is a major contribution toward reaching this goal? The answer to this question is the very topic of this book, and more specifically, of the present chapter. The goal is neither to draw a comprehensive review of everything known in this field, nor to provide the reader with a handbook of molecular epidemiology. It is rather to give my personal views on how molecular epidemiology should be conceived today and the future avenues this field may take.

3.2 How to Define Molecular Epidemiology?

Reading the literature, it is obvious that there is no consensus on how to define the term. The classical definition of the Centers for Disease Control (CDC) in Atlanta is "the various biochemical and molecular techniques used to type and subtype pathogens" (CDC 1994). The goal in this definition is very clear: identify the species (type) and subspecies/strains (subtype) of pathogenic organisms. The method is also very clear: the use of all techniques offered by biochemistry and molecular biology toward reaching this goal. In itself, this definition is quite satisfactory. The goal of molecular epidemiology is to help elucidate the routes traveled by epidemics by tracking the relevant units of analysis, the entities that are responsible for the epidemics. The difficulties lie in the very definition and delimitation of these entities, as we will see below.

It should be emphasized that molecular epidemiology is no more than a particular case of molecular systematics applied to the specific case of the identification of pathogenic agents. It would be highly desirable that the strict rules used to identify other organisms with molecular markers (Avise 2004) be applied to pathogens as well. As we will see, this is far from being the case.

3.3 Modern Technologies Have Upset the Field

Although there is no ultimate solution to all problems, there is no doubt that the arrival of powerful new technologies, and their lowering costs, have revolutionized the field. I will not detail here all the techniques that have been and are available for

pathogen characterization but rather explain how the techniques have made a major contribution to the field but are by no means a panacea.

In heroic times, the characterization of pathogenic agents and their strains relied on proteic markers, chiefly multilocus enzyme electrophoresis* (MLEE*, for a recent review see Tibayrenc 2009). These protein markers proved to be fine genetic markers and have allowed impressive progress, not only in routine strain typing, but also in basic bacterial population genetics (Selander et al. 1986). They have also been widely used for parasitic protozoa strain identification (Godfrey and Kilgour 1976). However, MLEE* has many drawbacks that definitely make it an outdated technology: (i) it is a time-consuming and delicate technique, (ii) it lacks resolution when the microevolutionary scale is considered, (iii) it is subject to homoplasy*, although this drawback has been exaggerated, (iv) and lastly, it requires bulk strain culturing, which leads to culturing bias* (see below).

Techniques that directly take DNA as the target molecule have definitely surpassed protein/MLEE* characterization. However, it should be stressed that the many valuable results gathered by MLEE* studies, especially when population genetics and the evolution of pathogens are considered, remain entirely valid and have been thoroughly confirmed by DNA techniques.

The major technological step that has upset the field of molecular epidemiology is polymerase chain reaction* (PCR). Indeed, PCR* makes it virtually possible to amplify the DNA of a single cell. Its potential resolution is therefore very high. Moreover, by using specific primers*, one can selectively amplify the DNA of the organism under study, even if it is mixed with foreign DNAs. These two properties obviate the need for bulk culturing, thus avoiding the culturing bias*.

Many different techniques rely on PCR* amplification, including random primed amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP). However, it can be said that two techniques presently dominate the field of molecular epidemiology, namely microsatellites* and multilocus sequence typing* (MLST). The main advantage of microsatellites* is their high resolution. MLST* has the advantage of its high standardization and perfect portability. In terms of resolution power, MLST* is not at the top: it is often surpassed by mere restriction fragment length polymorphism* (RFLP) of antigen genes. Lastly, thanks to the recent progress in automatic sequencing, whole genome sequencing (WGS) of many strains in bacterial species tends to become routinely accessible, and makes it possible to design many SNP (single nucleotide polymorphisms) to type strains of pathogens (Pearson et al. 2009) It should be stressed that in and of themselves, these techniques are mere labeling tools and say nothing about the biological and evolutionary properties of the organisms they aim to characterize (see below).

3.4 Has Molecular Epidemiology Helped Clinical Practice?

The answer to this question depends on what molecular epidemiology is considered to be. Strictly speaking, if we take the basic definition given by the CDC (1994), routine serological tests can be considered molecular epidemiologic tools. The same

holds true for PCR* identification of pathogen species. If we limit ourselves to these very basic tools, there is no doubt that molecular epidemiology has made a considerable contribution to routine medical practice. However, when a more common meaning is considered (that is to say, high-resolution identification of strains and clones), molecular epidemiology has not entered the routine daily practice of medical doctors (Humphreys 2004). It should instead be considered an epidemiological research tool. Still, the fact remains that it is not confined to basic research. It does help health professionals make practical decisions. The survey and control of the H5N1 epidemics, for example, would not have been possible without the contribution of molecular epidemiology.

Molecular epidemiology is at the very boundary between basic research, which might remain speculative, and operational biomedical research, in which sufficiently robust results should be unfailingly sought, because decisions need to be made.

3.5 Can Molecular Epidemiology Be Conceived of Without the Evolutionary Concepts?

This is apparently possible, since molecular epidemiology articles are published in high-impact journals that do not contain a word of evolutionary biology or population genetics (Foxman 2007; Sintchenko et al. 2007). Usually these articles rely on the implicit working hypothesis that natural populations of bacteria are composed of clones that undergo no genetic modifications over time. We will see that this is often untrue.

Logically speaking, evolutionary concepts should appear to be indispensable to analyze molecular data. The molecular polymorphism of organisms has been shaped by evolution. It is therefore rational to interpret it in terms of evolutionary genetics. When plants, mammals, insects, reptiles, etc. are investigated, it would appear nonsensical not to use population genetics and evolutionary biology concepts to analyze their molecular diversity (Avise 2004). Pathogens should not constitute a special case. Quite the opposite, I would say, because their formal genetics and mating strategies still remain mysterious, although considerable progress has been achieved in the last 20 years.

For years I have called for the inclusion of population genetics and evolutionary biology in the very definition of molecular epidemiology (Tibayrenc 1995, 2005).

According to this vision, molecular epidemiology should not limit itself to lazy band counting and mere dendrograms automatically generated by appropriate software. It should rely on classical population genetics (de Meeûs et al. 2007) and phylogenetic approaches (Hall and Barlow 2006).

This vision advocates molecular epidemiology as an exploration, using evolutionary biology's conceptual tools, of how pathogen populations are distributed and how they evolve.

3.6 Downstream Research

Additionally, I have long proposed that beyond this indispensable inventory, the consequences of pathogens' genetic diversity on their relevant biomedical properties (a disease's clinical diversity, sensitivity to drugs and vaccines, serological and molecular diagnosis) be explored (downstream studies, see Tibayrenc 1995). The minimum would be, for example, that a new drug or a new vaccine be tested on a set of selected strains that are representative of the entire genetic diversity of the species. With micropathogens, if technically feasible, it appears indispensable to work with laboratory-cloned stocks, with verification of the cloning under the microscope. Too many studies rely on noncloned stocks that are often composed of various different genotypes. This makes these mixtures of stocks highly unstable, and hence the results of such studies are poorly reproducible.

3.7 Reticulate Evolution of Pathogens and Its Implications

In brief, this is the main result yielded by evolutionary studies dealing with pathogens. This chapter is not the place for a long discourse on the evolutionary biology of microbes. To make a long story short, this result has been reached with two complementary means: population genetic statistics and phylogenetic analysis.

The term "reticulate evolution" by itself refers to the situation where separate evolution of distinct genetic lines is occasionally countered by limited genetic exchange. This situation is found in many plant species (Avise 2004).

Pathogenic microbes have evidenced a similar pattern in many, if not most, species (Tibayrenc et al. 1990, 1991; Awadalla 2003; Heitman 2006). Unexpectedly, many micropathogens exhibit sex in the broad sense of the term (i.e., exchange of genetic material between different cells, whatever the cytological mechanism involved). Its precise mechanisms vary considerably: conjugation, transduction and transformation in bacteria, classical meiosis in *Trypanosoma brucei*, the agent of human African trypanosomiasis (Jenni et al. 1986), and nonmeiotic hybridization in *Trypanosoma cruzi*, the agent of Chagas disease (Gaunt et al. 2003). However, the results in evolutionary terms are the same: departures from panmictic* expectation and from a purely phylogenetic view.

Departures from panmictic* expectations can be evaluated with classical segregation tests relying on Hardy-Weinberg and F-statistics (de Meeûs et al. 2007). However, these tests are based on the assumption that the organism under survey is diploid. This makes the use of these approaches impossible for haploid organisms (bacteria, the blood stage of *Plasmodium*) and debatable for those organisms for which diploidy is not fully ascertained (*Leishmania*, *Trypanosoma*). Recombination tests (linkage disequilibrium* analysis, Tibayrenc et al. 1990; Maynard Smith et al. 1993) avoid this shortcoming and are applicable whatever the ploidy of the organism may be, even if it is unknown, or if the molecular profiles make it impossible to

identify individual alleles (Tibayrenc et al. 1991). A considerable advantage of recombination tests is that, by definition, they are the only ones able to test for the stability of multilocus associations. This is the very goal of molecular epidemiology, as we will emphasize later.

Phylogenetic analysis, with its many different approaches and specialized softwares, looks for the presence and age of strictly separated genetic units. Its very definition makes departures from such an ideal image in themselves an indication of genetic exchange and its intensity. With highly recombining pathogens such as *Nesseiria gonorrhoeae* and *Helicobacter pylori*, any phylogenetic analysis is virtually impossible. Strong evidence for rarity or absence of recombination is a fair agreement between phylogenetic trees designed from different genetic markers, as seen in *T. cruzi* (see Fig. 3.1). This is a particular case of the concordance principle, which states that if a hypothesis is valid, it is increasingly confirmed by accumulating evidence (Avise 2004).

Pathogens show a continuum between these two extreme pictures, from highly recombining (*Helicobacter*, *Neisseria*) to scarcely recombining (*T. cruzi*, *Salmonella*, *Escherichia coli*).

The implications of this for molecular epidemiology and downstream studies are far-reaching. The more the species under study is composed of discrete, stable genetic lines (structured species, Tibayrenc 1995), the more these lines will be convenient targets for epidemiological follow-up and downstream studies. On the contrary, if genetic recombination is frequent, such discrete lines will not be present, and typing of stable multilocus genotypes will be impossible. In this case, only identification and follow-up of individual genes will be possible (see below).

3.8 Units of Identification: Genes, Clones, Strains, Subspecies, and Species

Founding molecular epidemiology on firm evolutionary concepts requires clearly defining the units of analysis to be used (Dijkshoorn et al. 2000).

The least questionable of these units of analysis is the gene. Many molecular epidemiology studies are based on the identification and follow-up of genes of interest (for example, genes of virulence, of drug resistance), conveniently labeled by specific PCR primers*. Such studies are possible whatever the population structure and recombination level of the pathogen under survey may be. However, it is more propitious to base these studies on a convenient and thorough population genetics framework of the species under study.

"Clone" should be understood here in its strict genetic meaning, that is to say, durable multilocus association. With this sense, a "clonal" species refers to all cases where the offspring genotypes are identical or almost identical to the parental ones, and where the propagation of multilocus genotypes that are stable is space and time are observed. This definition includes not only species that propagate themselves by classical mitotic propagation (the case of most bacterial species), but also many

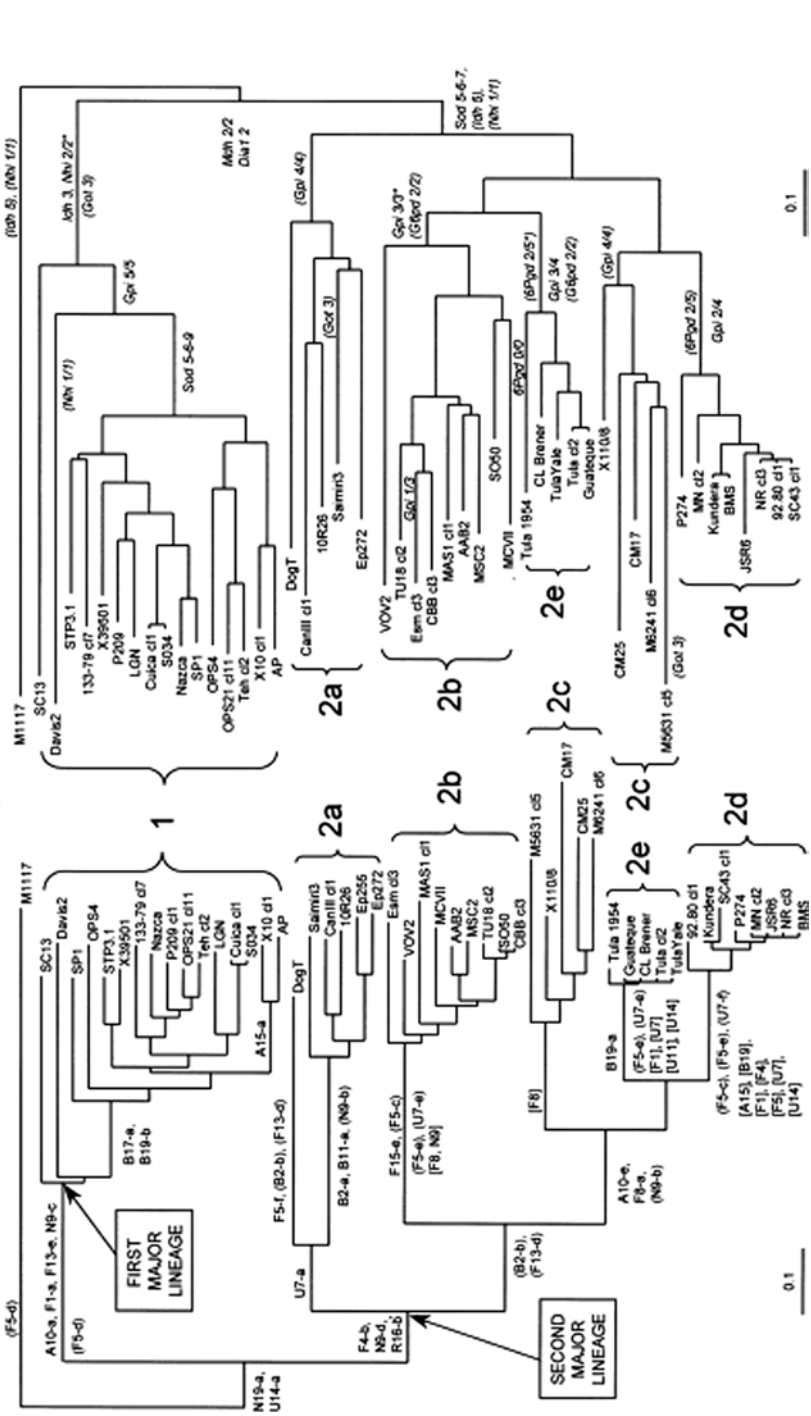

Fig. 3.1 Two phylogenetic trees depicting the evolutionary relationships among *T. cruzi* genotypes: isoenzymes (*left*) and RAPD (*right*). The strong similarity between the two trees is an extreme manifestation of linkage disequilibrium (After Brisse et al. 2000)

cases of parthenogenesis*, homogamy*, and self-fertilization* in homozygotes. Extreme inbreeding in this sense is not an alternative model to clonality (Rougeron et al. 2009), but rather a particular case of it (Tibayrenc et al. 2010).

All species in which genetic recombination is either absent or severely restricted should be considered clonal according to this definition, so that durable multilocus associations are not disrupted by recombination and behave like frozen genotypes or genetic photocopies. An important parameter to consider is the lifetime of such multilocus genotypes. Maynard Smith et al. (1993) distinguished between epidemic clonality and clonal evolution. The first refers to those species in which occasional bursts of genetic clones soon vanish in the common gene pool of a recombining species. *Neisseria meningitidis* is an example of this. Clonal evolution concerns those species in which clones remain stable at an evolutionary scale. *Trypanosoma cruzi* and *Salmonella* sp. are examples of this. Obviously, the consequences of the two models for molecular epidemiology and downstream studies are drastically different.

The term "strain" is widely used in the microbiology literature. However, its definition is a delicate one. Many specialists (including myself) use it with the sense of "stock", that is to say, a culture of a pathogen with a given origin. The World Health Organization's definition is quite different: it designates a set of stocks with various origins, but that share common defined properties (virulence, etc.). For molecular epidemiology, "strain" should designate a precise multilocus genotype, which is assumed to have a common, recent ancestry, in other words, a genetic clone. It is relevant to note that strains cannot exist in nonclonal, recombining species, although many epidemiologists maintain intuitively that such species have strains as well.

The definition of species is a never-ending story. An innumerable number of different definitions have been proposed by many authors (Hey 2001). However, one can distinguish between four main concepts to define species.

1. The biological species concept is the most classical one and refers to those communities of organisms that are potentially interfertile (Dobzhansky 1937). Obviously, this concept refers to classical, sexual species and is hardly useable for pathogenic microorganisms.
2. The phenetic species concept is based on the assumption that organisms pertaining to the same species are assumed to share many properties, while the contrary applies to organisms that belong to different species. To define species with this approach, as many characters as possible are taken into account without prioritizing them (Sneath and Sokal 1973). This concept could be useful to define pathogen species, although this approach is time-consuming.
3. The phenotypic concept is still widely used to define new species in insects, mammals, etc. In such cases, it is mainly based on morphological characters (color, shape, etc.). In microbiology, it can be said that many microbe species have been defined on phenotypic characteristics, either epidemiological traits or pathological properties. For example, *Mycobacterium leprae* is the causative agent of leprosy. *Leishmania infantum* is the agent of leishmaniosis in infants. *Leishmania panamensis* was first described in Panama. The phenotypic species concept applied to pathogenic microbes therefore aims at conveying relevant medical information.

4. The phylogenetic species concept considers that species should be equated to clades* or individual phylogenetic lines (Cracraft 1983). For the above-described reasons, this approach is not easy to use when micropathogens are considered due to the predominance of reticulate evolution in them. By definition, clades* do not exchange genes.

The definition of species for pathogens is a brain teaser. However, the extreme attitude that species are impossible to define in the case of microbes (de Meeûs et al. 2003) should be rejected. Health professionals need operational units of analysis to work on. Microbe species are not mere fantasies or video games: *Trypanosoma cruzi* cannot cause sleeping sickness. Even when considering the problem of reticulate evolution, one cannot see the microbial world as a genetic continuum. There are profound discontinuities, even if their boundaries are sometimes difficult to delineate. This means that the phylogenetic species concept can be used with some caution. The pragmatic approach I have recommended (Tibayrenc 2006) is to use the phylogenetic and phenotypic concepts jointly to describe new pathogen species. New species should be described only when they correspond to a phylogenetic reality and feature relevant medical/epidemiological properties. The use of the phylogenetic species concept means that species can be identified using the tools of molecular epidemiology.

In classical zoology, subspecies designate geographical variants of a given species that can be easily identified by specific traits, generally morphological particularities. The concept is of little value in medical microbiology. There are no strict criteria to define microbial subspecies; the tendency is for authors to describe subspecies when they dare not make real new species (Schönian et al. 2010).

3.9 Helpful, Although Not So Successful, Operational Concepts: Clonets and Discrete Typing Units/Tags

The classical concepts of population genetics and phylogenetic analysis are sometimes poorly adapted to the peculiarities of the microbial world and the specific demands of epidemiological follow-up.

3.9.1 Clones and Clonets

Let us consider a species that has been evidenced to be clonal by population genetic analysis. The use of a given set of genetic markers, say 20 microsatellite* loci, has individualized stocks that share the same multilocus combination. Do they represent genuine clones? Probably not. If we use 30 microsatellite* loci instead of 20, we will probably encompass additional genetic variability within each of these supposed

clones. To get around the difficulty, the term "clonet" has been coined to refer to those sets of stocks that appear to be identical to a given set of genetic markers in a basically clonal* species (Tibayrenc and Ayala 1991). The clonets are relevant units of analysis for epidemiological tracking. They should be delimited by a sufficient range of genetic markers, according to the time and space scale considered (see below).

3.9.2 Discrete Typing Units and Tags

I have already presented the implication of reticulate evolution for phylogenetic/cladistic* analysis. The presence of some gene flow among genetic lines, and the fact that in hybridization events some genetic lines have two ancestors instead of one, make the clade* concept invalid. There are obvious genetic discontinuities between and within pathogen species that deserve to be described and delimited. To avoid the problem, I have proposed the descriptive and operational concept of discrete typing unit (DTU) (Tibayrenc 1998). DTUs are sets of stocks that are genetically closer to each other than to any other stock and are identifiable by common molecular, genetic, biochemical, or immunological markers called tags. DTUs correspond to reliable units of analysis for molecular epidemiology and downstream studies.

3.10 Can This Field Be Unified?

People designing and using molecular epidemiology sorely need to realize that they have common goals and that their approaches should converge to a large extent. Whatever the pathogen under survey – parasite, fungus, bacterium, virus, prion, be it of medical, veterinary or agronomical relevance – species, strains, clones, and genes of interest should be reliably identified using evolutionary concepts. The genetic variability of the pathogen considered should be taken into account for vaccine/drug design, epidemiological follow-up and clinical studies.

Yet the world of pathogens is too heterogeneous for building totally standardized approaches. Species concepts are not the same for parasites, bacteria, and viruses. What's more, the molecular tools cannot be the same.

However, when the pathogens considered are closely related, it is nonsense to design radically different approaches. An example is the description of new species. With *Leishmania* parasites, many new species have been described, some of them based on doubtful criteria (Van der Auwera et al. 2011). On the other hand, until now, specialists of Chagas disease have agreed on considering *Trypanosoma cruzi*, the causative agent of Chagas disease, a unique, although genetically extremely heterogeneous, species (Zingales et al. 2009). *Leishmania* specialists are splitters, Chagas specialists are lumpers. These differences have little evolutionary or medical justification.

3.11 The Crucial Parameters of Time and Space Scales

Molecular epidemiology may act on very different time and space scales. For example, one may wish to survey the spread of one of the six *T. cruzi* DTUs over the entire American continent. A totally different case would be the identification and follow-up of a unique methicillin-resistant clone of *Staphylococcus aureus* that has contaminated a hospital intensive care unit. Obviously the molecular and statistical tools will have to be adapted to the case at hand. In the first case, classical phylogenetic analysis and markers that have a slow molecular clock* (multilocus sequence typing*, multilocus enzyme electrophoresis*) can be used. In the second case, markers with a finer resolution, in other words, a faster molecular clock*, such as microsatellites* or some RFLP* markers, will have to be used.

3.12 New Problems: Typing Noncultivable Pathogens

Koch's postulates state that:

(a) The microorganism must be found in abundance in all organisms suffering from the disease but should not be found in healthy animals.
(b) The microorganism must be isolated from a diseased organism and grown in pure culture.
(c) The cultured microorganism should cause disease when introduced into a healthy organism.
(d) The microorganism must be reisolated from the inoculated, diseased experimental host and identified as being identical to the original specific causative agent.

These postulates are now invalidated by the fact that many pathologies, such as Crohn disease, could be caused by pathogens that cannot be cultured.

One of the most remarkable achievements of molecular epidemiology is to make it possible to characterize these unknown pathogenic agents. Sequences of nonspecific genes such as ribosomal RNA genes are used to reach this goal (see Frank and Feldman 2007 for a recent, exhaustive review).

3.13 Conclusion: The Future of the Field: Total Pathogen Profiling, Whole Genome Sequencing, Integrated Genetic Epidemiology

Technological progress will continue at an increased speed, just as the cost of analyses will continue to be lowered. This will permit holistic approaches that were not possible with limited technological resources. Pathogen profiling is one of the new strategies that will emerge from use of these new tools

(Sintchenko et al. 2007). It integrates molecular data (genome, transcriptome, proteome, metabolome) with clinical and epidemiological data, assisted by geographic information systems (GIS).

Pathogen profiling will be even more valuable if it relies on the survey of pathogen's whole genomes, which becomes now more and more accessible, at least for viruses and bacteria (Pearson et al. 2009). This opens the way to the new field of population genomics (Tibayrenc 2005).

Another approach long called for (Tibayrenc 1998) is to concurrently consider the role played in the transmission and severity of infectious diseases, by the host's, the pathogen's and the vector's (in the case of vector-borne diseases) genetic diversity as well as the interactions between the three (coevolution phenomena). Such a holistic approach (integrated genetic epidemiology of infectious diseases) should be extended to the entire genome of hosts, pathogens and vector (population genomics). This is the very topic of the new journal Infection, Genetics and Evolution (Elsevier; http://www.elsevier.com/locate/meegid) and the Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases (MEEGID; http://www.meegidconference.com/) congresses. Such an approach has been proposed for the study of Chagas disease (Tibayrenc et al. 2010).

Glossary of Specialized Terms (* in text)

Clade evolutionary line defined by cladistic analysis. A clade is monophyletic* and is genetically separated (i.e., evolves independently) from other clades.

Cladistic analysis a specific approach of phylogenetic analysis based on the polarization of characters that are shared between ancestral (plesiomorphic) and derived (apomorphic) characteristics. Only those apomorphic characters shared by all members of a given clade* (synapomorphic character) are considered to convey relevant phylogenetic information (Hennig 1966).

Culturing bias This term refers to the situation where a host is infected by several different genotypes of a given pathogen. When culturing an isolate of this pathogen, some genotypes will tend to dominate to the detriment of others. At the end of the culturing process, the collection of genotypes does not reflect the original isolate (Tibayrenc 1995).

Homogamy preferential mating between individuals that are genetically identical or extremely similar to each other.

Homoplasy Common possession by distinct evolutionary units of identical characters that do not originate from a common ancestor. The origin of homoplastic traits include the following: (a) convergence (possession of identical characters derived from different ancestral characters, due to convergent evolutionary pressure, for example, the fins of fish and dolphins); (b) parallelism (possession of identical characters derived from a single ancestral character and generated independently in different evolutionary units); and (c) reversion (restoration of an ancestral character from a derived character).

Isoenzymes Protein extracts of the biological samples under analysis are separated by electrophoresis. The gel is then processed with a histochemical reaction involving the specific substrate of a given enzyme. This enzyme's zone of activity is then specifically stained on the gel. From one sample to another, migration differences can be visible for this same enzyme. These different electrophoretic forms of a given enzyme are referred to as isoenzymes or isozymes. When given isoenzymes are driven by different alleles of a single gene, they are more specifically referred to as alloenzymes or allozymes. Differences in migration result from different overall electrical charges between isoenzymes. Overall electric charges are a resultant of the individual electric charges of each amino acid (AA) of a given enzyme. The AA sequence is the direct result of the DNA sequence of the gene that codes for this enzyme. It is therefore considered (and verified) that isoenzyme polymorphism is a faithful reflection of the genetic polymorphism of the organism under study.

Linkage disequilibrium Nonrandom association of genotypes occurring at different loci.

Microsatellite A short DNA sequence of DNA, usually 1–4 bp long, that is repeated together in a row along the DNA molecule. Microsatellites are fast-evolving markers, with a high resolution level and are found in many organisms, including pathogens.

Molecular clock In its strict, original sense (more correctly called the DNA clock hypothesis), the concept that the rate of nucleotide substitutions in DNA remains constant over time. In a broader sense, simply how fast the genomic part that codes for the variability of the marker considered evolves. This speed is driven by the mutation rate. It may be regular or irregular.

Monophyletic an evolutionary line that has a unique ancestor.

Multilocus sequence typing (Maiden et al. 1998) is a highly standardized approach based on the sequencing of 450-bp parts of a set of housekeeping genes (usually seven). It has been widely used for a high number of bacterial species and some eukaryotic pathogens as well. The main advantage of MLST is it perfect portability (possibility of reliably communicating results among different laboratories), since sequences can be simply emailed. Strains that share the same combination of alleles are referred to as sequence types (ST). Strains that share 7/7 alleles = consensus group; strains that share 6/7 alleles = single-locus variants (SLV); strains that share 5/7 alleles are double-locus variants (DLV). SLV and DLV = clonal complexes.

Panmixia panmictic: Situation in which genetic exchanges occur randomly in the population under survey.

Parthenogenesis Reproduction by the development of a single gamete without fertilization by a gamete of the opposite sex.

Polymerase chain reaction (PCR) A technique that copies the complementary strands of a target DNA chain through a set of cycles until the needed DNA amount is produced. PCR uses synthesized primers* whose nucleotide sequences are complementary to the DNA flanking the target region. The DNA is heated to separate the complementary strands, then cooled to have the primers bind

to the flanking sequences. The enzyme Taq DNA polymerase is added and the reaction is left to pass through the required number of replication cycles.

Primer a short DNA sequence used in PCR* technologies, that anneals to a single strand of DNA and acts as a starting point to initiate DNA polymerization mediated by the enzyme Taq DNA polymerase.

Restriction fragment length polymorphism (RFLP) variability in the DNA of a given organism evidenced by the use of restriction endonuclease bacterial enzymes. The endonuclease cuts the DNA at a specific restriction site with a given sequence, and the polymorphism of the DNA fragments thus obtained can be visualized on gels, either directly by ethidium bromide staining or by Southern blot hybridization with specific probes.

Self-fertilization fertilization by the union of male and female gametes from the same individual

References

Avise JC (2004) Molecular markers, natural history and evolution, 2nd edn. Chapman & Hall, New York

Awadalla P (2003) The evolutionary genomics of pathogen recombination. Nat Rev Genet 4:50–60

Brisse S, Barnabé C, Tibayrenc M (2000) Identification of six *Trypanosoma cruzi* phylogenetic lineages by random amplified polymorphic DNA and multilocus enzyme electrophoresis. Int J Parasitol 30:35–44

Centers for Disease Control and Prevention (1994) Addressing emerging infectious diseases threat. A prevention strategy for the United States, 27, Atlanta

Cracraft J (1983) Species concept and speciation analysis. In: Johnson RF (ed) Current ornithology. Plenum Press, New York, pp 159–187

de Meeûs T, Durand P, Renaud F (2003) Species concepts: what for? Trends Parasitol 19:425–427

de Meeûs T, McCoy K, Prugnolle F et al (2007) Population genetics and molecular epidemiology or how to "débusquer la bête". Infect Genet Evol 7:308–332

Dijkshoorn L, Ursing BM, Ursing JB (2000) Strain, clone and species: comments on three basic concepts of bacteriology. J Med Microbiol 40:397–401

Dobzhansky T (1937) Genetics and the origin of species. Columbia University Press, New York

Foxman B (2007) Contributions of molecular epidemiology to the understanding of infectious disease transmission, pathogenesis, and evolution. Ann Epidemiol 17:148–156

Frank DN, Feldman RA (2007) Molecular-phylogenetic strategies for characterization of uncultured pathogens. In: Tibayrenc M (ed) Encyclopedia of infectious diseases- modern methodologies. Wiley, New York, pp 57–73

Gaunt MW, Yeo M, Frame IA et al (2003) Mechanism of genetic exchange in American trypanosomes. Nature 421:936–939

Godfrey DG, Kilgour V (1976) Enzyme electrophoresis in characterizing the causative agent of Gambian Trypanosomiasis. T Roy Soc Trop Med H 71:217–225

Haldane JBS (1949) Disease and evolution. La ricerca scientifica 19(Suppl):68–76

Hall BG, Barlow M (2006) Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. Ann Epidemiol 16:157–169

Heitman J (2006) Sexual reproduction and the evolution of microbial pathogens. Current Biol 16:R711–R725

Hennig W (1966) Phylogenetic systematics. Univ Illinois Press, Urbana

Hey J (2001) The mind of the species problem. Trends Ecol Evol 16:326–329

Humphreys H (2004) Does molecular typing make any contribution to the care of patients with infection? Clin Microbiol Infect 10:269–271

Jenni L, Marti S, Schweizer J et al (1986) Hybrid formation between African trypanosomes during cyclical transmission. Nature 322:173–175

Maiden MCJ, Bygraves JA, Feil E et al (1998) Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA 95:3140–3145

Maynard Smith J, Smith NH, O'Rourke M et al (1993) How clonal are bacteria? Proc Natl Acad Sci USA 90:4384–4388

Pearson P, Richard T, Okinaka RT et al (2009) Phylogenetic understanding of clonal populations in an era of whole genome sequencing. Infect Genet Evol 9:1010–1019

Rougeron V, De Meeûs T, Hide M et al (2009) Extreme inbreeding in *Leishmania braziliensis*. Proc Natl Acad Sci USA 25:10224–10229

Schönian G, Mauricio I, Cupolillo E (2010) Is it time to revise the nomenclature of *Leishmania*? Trends Parasitol 26:466–469

Selander RK, Musser JM, Caugant DA et al (1986) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol 51:873–884

Sintchenko V, Iredell JR, Gilbert GL (2007) Pathogen profiling for disease management and surveillance. Nature Rev Microbiol 5:464–470

Sneath PHA, Sokal RR (1973) Numerical taxonomy. In: Kennedy D, Park RB (eds) The principle and practice of numerical classification. Freeman, San Francisco, pp 537

Tibayrenc M (1995) Population genetics of parasitic protozoa and other microorganisms. Adv Parasitol 36:47–115

Tibayrenc M (1998) Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. Int J Parasitol 28:85–104

Tibayrenc M (2001) The golden age of genetics and the dark age of infectious diseases. Infect Genet Evol 1:1–2

Tibayrenc M (2005) Bridging up the gap between molecular epidemiologists and evolutionists. Trends Microbiol 13:575–580

Tibayrenc M (2006) The species concept in parasites and other pathogens: a pragmatic approach? Trends Parasitol 22:66–70

Tibayrenc M (2009) Multilocus enzyme electrophoresis for parasites and other pathogens. In: Caugant DA (ed) Methods in molecular biology: molecular epidemiology of microorganisms. Humana Press, Totowa, pp 13–25

Tibayrenc M, Ayala FJ (1991) Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. Parasitol Today 7:228–232

Tibayrenc M, Kjellberg F, Ayala FJ (1990) A clonal theory of parasitic protozoa: the population structure of Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas and Trypanosoma, and its medical and taxonomical consequences. Proc Natl Acad Sci USA 87:2414–2418

Tibayrenc M, Kjellberg F, Arnaud J et al (1991) Are eucaryotic microorganisms clonal or sexual? A population genetics vantage. Proc Natl Acad Sci USA 88:5129–5133

Tibayrenc M, Telleria J, Diosque P et al (2010) Integrated genetic epidemiology of Chagas disease. In: Tibayrenc M (ed) Genetics and evolution of infectious diseases. Elsevier Insights, Burlington

Van der Auwera G, Fraga J, Montalvo AM et al (2011) *Leishmania* taxonomy up for promotion? Trends Parasitol 27:49–50

Zingales B, Andrade SG, Briones MRS et al (2009) A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. Mem Inst Oswaldo Cruz 104:1051–1054

Chapter 4
Population Genetics and Molecular Epidemiology of Infectious Diseases

Christine Chevillon, Thierry de Meeûs, and Karen D. McCoy

Abstract Population genetics is based on analysing the polymorphism patterns of genetic markers at different organizational levels; *i.e.* within and between individuals sampled among populations. From such analyses, inferences can be made on the reproductive modes of the species in question (hence, on patterns of genetic transmission along successive generations), as well as on the demographic functioning of the studied populations (i.e., population sizes, dispersal rates among populations, etc.). In this chapter, we present the main bases of population genetics theory and illustrate its interest for epidemiological issues via different case studies on parasite species.

4.1 Introduction

Due to their small size, location, biology and behaviour, direct observations on the population biology of parasites (e.g., dispersal ranges, population sizes or variation in population sizes) are almost impossible. However, molecular markers and computational tools from population genetics and phylogenetics can provide some of

C. Chevillon (✉) • K.D. McCoy
Maladies Infectieuses & Vecteurs: Ecologie, Génétique, Evolution & Contrôle (MIVEGEC), UMR 5290 CNRS-IRD-Universités UM1 & UM2, Centre IRD, 911 avenue, Agropolis, BP 64501, F-34000 Montpellier, France
e-mail: christine.chevillon@ird.fr; karen.mccoy@ird.fr

T. de Meeûs
Interactions hôtes-vecteurs-parasites dans les infections par des trypanosomatidae (TRYPANOSOM), UMR IRD/CIRAD 177, Centre International de Recherche-Développement sur l'Elevage en zone Subhumide (CIRDES), 01 BP 454, Bobo-Dioulasso 01, Burkina-Faso, Africa

CNRS, Délégation Languedoc-Roussillon, 1919, route de Mende – 34293, Montpellier Cedex 5, France
e-mail: thierry.demeeus@ird.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_4,

this essential information and can be used to track parasite dispersal at various scales (i.e., among host individuals, host populations and host species), thus reconstructing a parasite's epidemiological history (Criscione et al. 2005; de Meeûs et al. 2007; Archie et al. 2009). Phylogenetic analyses can help characterise the evolution and demography of rapidly evolving parasites such as viruses (e.g., for case studies and computational tools see (Pybus et al. 2001, 2003; Real et al. 2005; Drummond and Rambaut 2007)), and can even identify and characterise recombination events within datasets that include both parent and offspring sequences (Martin et al. 2005a, b). However, for making inferences on the migration, demography and reproductive modes and/or strategies of more slowly evolving parasites, population genetics tools are more adapted than phylogenetic ones. This chapter will mostly focus on cases for which population genetics tools are the most powerful; i.e., for analysing the polymorphism of nuclear markers to understand the reproduction, dispersal and demographic functioning of diploid Eukaryote species. In such species, each individual bears two alleles transmitted via two gametes, with these gametes having been produced either by two parents or by a unique hermaphroditic organism.

Understanding how population genetics inferences can be derived from patterns of genetic polymorphisms within and among parasite-individuals first requires that we characterise the action of several evolutionary forces on these patterns. To do so, population genetics theory has defined a pivotal null hypothesis: the so-called Hardy-Weinberg equilibrium (HWE) that describes the ideal situation of a non-evolving population. In the first section of this chapter, we briefly revisit this null hypothesis along with the most interesting aspects of population genetics from an epidemiological perspective. In particular, we discuss the conditions required for a population to be at HWE, the main characteristics of the evolutionary forces driving populations away from HWE, and the way that F-statistics can be used to characterise parasite transmission among hosts and genetic transmission along parasite generations. We will then say a few words on selection and present a case study where population genetics was used to identify the genetic basis of parasite drug resistance. The last two sections focus on the analysis of non-selected polymorphisms aimed at characterising (i) the transmission patterns of parasites at various geographical scales, and (ii) the transmission patterns of genetic polymorphisms along parasite generations (i.e., reproductive modes).

4.2 A Few Definitions and an Introduction to Population Genetics Tools

Our aim here is far from a detailed presentation of all the population genetics tools currently available, but rather to recall the basic notions required to understand the inferences presented in the case studies. The reader can find such a detailed (yet easily accessible) presentation of almost all existing analytical tools, as well as their conceptual bases and the power and limits of the associated statistical tests in de

Meeûs et al. (2007). Further information on population genetics theory can be found in basic text books such as those by Hartl and Clarke (1989) or Hedrick (2005). More experienced readers would also benefit from Rousset (2004).

4.2.1 Hardy-Weinberg Equilibrium

HWE (Hardy-Weinberg equilibrium) refers to an ideal state of a population within which the polymorphism of a nuclear marker of a diploid species within a population remains identical across generations (i.e., identical allele frequencies AND identical frequencies of all homozygous and heterozygous genotypes). This equilibrium exists, if and only if, the following assumptions are fulfilled:

- There is no mutation
- There is no selection (i.e., there is an even distribution of all individuals to the next generation, with thus a random sampling of all alleles present in the parental generation to form offspring)
- The studied population is of infinite size
- The studied population is isolated (i.e., there is neither immigration into the population nor emigration from it)
- Reproduction is sexual with a random association of gametes produced in the population (i.e., panmixia*)

Under HWE, the genotypic frequencies are entirely determined by the allelic frequencies. Let's consider a molecular marker with two alleles *A1* and *A2* present at frequencies p and q = 1–p, respectively in the population. The frequency of *A1A1* homozygous individuals will be given by the probability of sampling the *A1* allele twice within the population, thus p^2. The frequency of *A2A2* homozygous individuals will be determined by the probability of sampling the *A2* allele twice, thus q^2. The frequency of *A1A2* heterozygous individuals will be given by the probability of first sampling allele *A1,* then sampling the A2 allele, or of sampling them in the reverse order, thus $pq + qp = 2pq = 1 - p^2 - q^2$.

4.2.2 Micro-evolutionary Forces

The required conditions for HWE automatically define the evolutionary forces that affect allelic frequencies and/or the relationships among allelic and genotypic frequencies. Such evolutionary forces are genetic drift due to the finite sizes of real populations, migration of individuals among populations, mutation, selection, and non-random associations of gametes (i.e., deviation from panmixia).

Among these forces, the particularity of mutation is two-fold. On the one hand, it is the force that ultimately creates genetic polymorphism and hence enables a population to evolve. On the other hand, mutation rates are typically low enough

Table 4.1 Reproduction modes and impacts on deviation from HWE

Reproduction mode	Deviation from HWE: Target	Deviation from HWE: Impact on heterozygous frequency % HWE	Examples & references
Selfing	Whole genome	Deficit	*Tenia solium* is an obligatory selfing parasite with effective population size reduced by 1/2 (Kunz 2002; de Meeus et al. 2003)
Clonality	Whole genome	Excess	Many parasites including the fungus *Candida albicans* (see point 3–2.) See de Meeûs et al. (2006) for advices on sampling design and theory
Homogamy	Genes involved in the definition of the selected character (and physically linked loci)	Deficit	This implies that one prefers mating with a partner sharing the same phenotypic characteristics than itself, such as size for instance See point 3–1. for testing its occurrence
Heterogamy	Genes involved in the definition of the selected character (and physically linked loci)	Excess tending to be maximal	E.g. sexual locus in all species with genetically-defined sex See point 3–1. for testing its occurrence

than its influence can often be neglected in population genetic analyses. In contrast, the other evolutionary forces are often highly influential in population genetics. Real populations are of finite size and exchange migrants. As such, they all experience the combined action of genetic drift and migration. A finite population size means that, among all possible offspring that could be produced by a parental generation, only a sub-sample will occur and survive. This sampling effect, referred to as genetic drift, induces random fluctuations in allelic frequencies across generations until a point in time when, by chance (and in the absence of mutation), one allele will be fixed in the population. As genetic drift is a random effect, it reduces the probability that nearby populations display the same distribution of allelic frequencies (i.e., it increases the genetic differentiation among populations). Its effect is proportional to the effective size of the population; the smaller the population size, the greater the effect of drift on allele frequencies. The differentiating effect of genetic drift is counterbalanced by the action of migration which tends to homogenize allelic frequencies among populations exchanging migrants. Compared to mutation, migration and drift which typically affect the whole genome, selection can usually only be seen at a few associated loci (i.e., due to associated selection or direct linkage); selected loci oppose 'neutral loci' at which polymorphisms evolve without any consequences for the fitness of the organism. Finally, HWE can be disrupted due to an organism's reproductive mode. Indeed, there are many ways to deviate from random mating within populations; all of them modify the genotypic

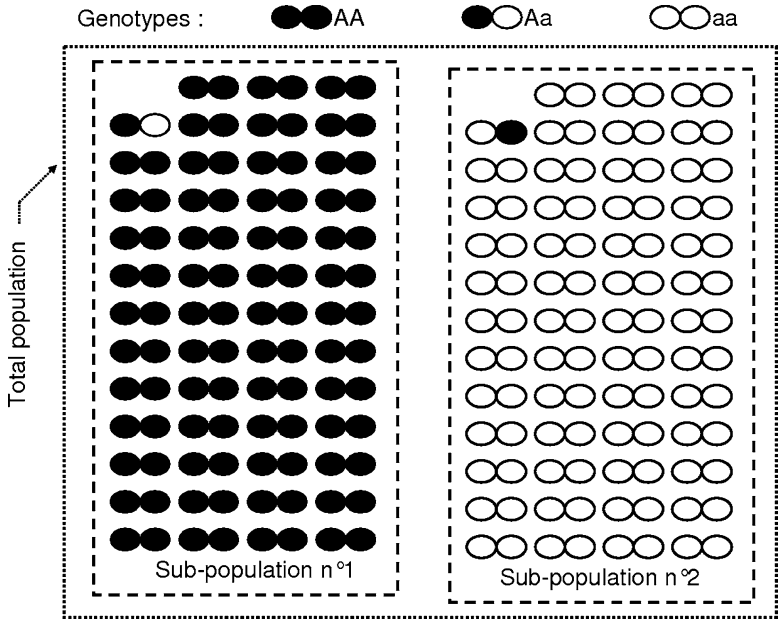

Fig. 4.1 Illustration of a Wahlund effect by analysing an extreme situation. Let us consider a population made up of two sub-populations of size N = 51 that are almost fixed for alternative alleles of a diallelic locus: each of the two populations groups 50 identical homozygous individuals and a single heterozygote. Such sub-populations do not deviate from HWE as the frequency of the almost fixed allele is given by $p = 101/102 \sim 0.99$, that of the rare allele is given by $q = 1/102 \sim 0.01$, and thus the number of expected heterozygous individual under HWE would be given by $2pqN \sim 1$. Things are very different for the total population. If we ignored the presence of two genetically differentiated sub-populations, we would observe a population of size N = 102 with identical frequencies of both alleles (p = q = ½). However, in this case, we only observe two heterozygous individuals rather than the 2pqN = 51 expected under HWE

composition of the population resulting in either an excess or a deficit of heterozygous individuals relative to HWE expectations at either one or all possible loci (see Table 4.1 for different reproductive modes and their influence on HWE).

4.2.3 Characterising Population Structure

Natural populations are usually organized such that individuals mostly reproduce within sub-populations of limited sizes, or demes (hence, experience genetic drift). Ignoring this population structure when computing allelic and genotypic frequencies – i.e., computing frequencies over the total population, rather than within its sub-populations where reproduction and demographic regulation take place- results in a heterozygote deficit relative to HWE, if subpopulations display different allelic frequencies. This is the so-called Wahlund effect (Fig. 4.1).

4.2.3.1 F-Statistics: Definition

Wright (1965)'s F-statistics allows one to disentangle the effects of population structure from those due to deviations from panmixia within sub-populations, and hence to characterise the demographic functioning of a subdivided population. F_{IS} reflects inbreeding due to deviations from panmixia within subpopulations. F_{ST} corresponds to the inbreeding caused by the subdivision of the total population into subpopulations of limited size. F_{IT} evaluates the inbreeding of individuals in the total population that results from both previous effects, local deviations from panmixia and population structure. Given Q_I, the average probability of identity* (in descent or in state) between two alleles sampled from the same individual, Q_S, the average probability of identity (in descent or in state) between two alleles sampled from different individuals of the same sub-population, and Q_T, the average probability of identity (in descent or in state) between two alleles sampled from individuals from two distinct sub-populations of the total population, the three F-statistics can then be defined as follows (Rousset 2004):

$$F_{IS} = \left(Q_I - Q_S\right)\Big/\left(1 - Q_S\right)$$

$$F_{ST} = \left(Q_S - Q_T\right)\Big/\left(1 - Q_T\right)$$

$$F_{IT} = \left(Q_I - Q_T\right)\Big/\left(1 - Q_T\right)$$

F_{IS} can take any value between −1 (all individuals are heterozygous for the two same alleles) and 1 (none of the individuals are heterozygous in polymorphic subpopulations), with the null value corresponding to the case where sub-populations are at HWE. F_{ST} can take any value between 0 (when genetic drift has no effect) and 1 (when all subpopulations have fixed alternative alleles). F_{IT} takes any value between −1 (when all individuals are heterozygous for the same two alleles across the total population) and 1 (when all subpopulations are polymorphic, but all individuals are homozygous).

To compute F-statistics, data on the distribution of allelic and genotypic frequencies across subpopulations are required. Congruency among loci using per-locus estimates of F-statistics is then examined to assess the joint effect of migration and genetic drift: indeed, by definition, these evolutionary forces are expected to apply evenly over the entire genome of the studied species! In other words, discrepancies in *F-statistics* estimates among loci can help identify genotyping problems and/or selected polymorphisms (Fig. 4.2).

Weir and Cockerham (1984) defined the respective unbiased estimators f, θ and F for the F_{IS}, F_{ST} and F_{IT} parameters. Population geneticists can obtain these unbiased estimators using software such as *GENEPOP* (Raymond and Rousset 1995) or *F-stat* (Goudet 1995). Excoffier et al. (1992) designed the Φ − statistics, the equivalent of

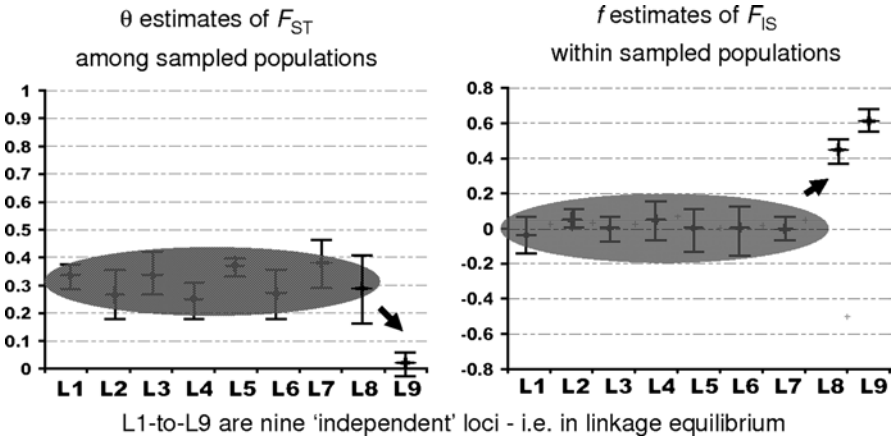

Fig. 4.2 Comparisons of locus-specific estimates of *F-statistics*. For each F-statistic estimate, population genetics software may compute confidence intervals. Here, we present a putative analysis based on the polymorphism of nine independent genetic markers (i.e., in linkage equilibrium). The means and standard errors of the per locus estimates of the parameters F_{ST} (on *left*, estimator θ) and F_{IS} (on *right*, estimator *f*) are displayed for the same set of populations for the nine independent loci. Seven loci provide congruent estimates for both parameters: they are thus likely to provide the information we are looking for, that is the population structure that results from the balance among genetic drift, migration and reproduction. Loci L8 and L9 provide different parameter estimates compared to the other seven and should thus be considered with caution. Relative to the consensus estimates provided by loci L1 to L7, loci L8 and L9 both display heterozygote deficits within populations. However, only locus L9 shows a drastic reduction in the estimate of F_{ST}. Based on these patterns, one could hypothesise that deviations in locus L8 may be due to the presence of a null allele, whereas estimates at L9 may be due to either the physical linkage of this locus to the sex-determining locus of a bisexual species (with all sampled populations tending toward an even sex-ratio) and/or to the action of diversifying selection within the sampled populations

F-statistics for haploid polymorphism (e.g., markers located in mitochondrial DNA). Population geneticists may use the *Arlequin* software (Excoffier et al. 2005) for computing unbiased estimates of the Φ – statistics.

4.2.3.2 Optimizing the Accuracy of F-Statistics Estimates

Of course, in order to obtain accurate estimates of F_{IS}, F_{ST} and F_{IT} and thus on the population structure of the organism of interest, we require genetic markers that unambiguously discriminate all possible heterozygous and homozygous genotypes from one another (i.e., so-called codominant markers). This is usually, although not obligatorily, the case for the most popular genetic markers used in population genetics studies; i.e., allozymes (also called isoenzymes) used at the end of twentieth century, microsatellite loci in use for the past two decades, and the promising SNPs (single nucleotide polymorphisms).

Allozymes are metabolic enzymes. The alleles which have conserved enzymatic activities are discriminated via their relative rapidity to migrate in an electrophoresis gel; such a migration rapidity depends on their differences in electric charge and molecular weight (see for details on enzyme electrophoresis Pasteur et al. 1988). The knowledge of the enzymatic properties of the allozyme of interest allows defining the substrate to furnish for obtaining a coloured product revealing the relative position of allelic enzymes. Such a genotyping process requires thus to have conserved samples since collection time under conditions that prevents the degradation of enzymes (e.g. in liquid nitrogen). Depending of the enzyme structure (monomeric, dimeric or tetrameric), heterozygous for alleles that have conserved their enzymatic properties will display from one to five bands, and hence be easily discriminated from either homozygous genotypes (Pasteur et al. 1988). However, in some cases, alleles which have lost their enzymatic properties are kept within populations; in such a case any heterozygous individual for such null alleles and a detectable allele will be mistaken for a homozygous for the detectable allele. For instance, heterozygous deficits were recurrently observed in allozyme-based population genetic studies of marine bivalves. David et al. (1997) showed that, in the *Spisula ovalis* species, the most likely explanation of such a phenomenon was that most of the allozymic loci surveyed displayed null alleles at low frequency.

Microsatellite loci are short tandem repeated of sequence motifs of two to five base pairs, such that the action of mutations at such loci is considered to mainly result in a variation in the repetition number of such motifs. The knowledge of the sequences that flank such repetitions at either side allows designing PCR-primers and -conditions to amplify a chromosomal region encompassing a microsatellite locus. From then, the electrophoresis of PCR-fragments allows scoring the one or two alleles of distinguishable sizes that constitute the individual genotype. With such a genotyping process, the only alleles that cannot be scored (null alleles) are those for which the flanking regions recognized by the PCR-primers have mutated. Again, a heterozygous individual for both null and non-null alleles will be confused with a homozygous for the non –null allele. Irrespectively to the nature of the marker used, the presence of null alleles lead thus to artificially increase the homozygous frequencies at the concerned locus and within the populations where the null alleles exist. As a result, loci at which null alleles exist within the studied samples will lead to different estimates of the F-statistics than the other loci. This is another reason, which is independent from the potential presence of selected polymorphisms among the chosen loci, to base population genetics analyses on a minimum of six-to-ten loci. The loci leading to similar estimates of *F-statistics* are thus likely to be neutral and null-allele free, and thus to precisely inform on the balance achieved among the sampled populations among migration, genetic drift and reproduction mode.

SNPs stand for single nucleotide polymorphisms; i.e., single base pair positions in genomic DNA which are polymorphic in state (i.e., without cases of insertions and/or deletions) among the sequences driven for individuals. SNPs are usually bi-allelic, rather than tetra-allelic as one might *a priori* expect (Brookes 1999). The usual way to define SNPs is sequencing and/or next-generation sequencing.

4.2.3.3 Linkage Disequilibrium

Linkage disequilibrium refers to the correlation between the polymorphism recorded at two loci. The converse, linkage equilibrium, then refers to the independence of the two loci. For the sake of simplicity, let us consider two bi-allelic loci A (with alleles *A1* and *A2*) and B (with alleles *B1* and *B2*) so that the genotypic frequencies recorded within a population are given, at locus A, by p_{A1A1}, p_{A1A2} and p_{A2A2}, and, at locus B by p_{B1B1}, p_{B1B2}, and p_{B2B2}. These two loci are in linkage equilibrium whenever the frequencies of the nine bi-locus genotypes are determined by the product of the two mono-locus genotypic frequencies:

$$p_{A1A1B1B1} = p_{A1A1} \times p_{B1B1} \quad p_{A1A2B1B1} = p_{A1A2} \times p_{B1B1} \quad p_{A2A2B1B1} = p_{A2A2} \times p_{B1B1}$$
$$p_{A1A1B1B2} = p_{A1A1} \times p_{B1B2} \quad p_{A1A2B1B2} = p_{A1A2} \times p_{B1B2} \quad p_{A2A2B1B2} = p_{A2A2} \times p_{B1B2}$$
$$p_{A1AB2B2} = p_{A1A1} \times p_{B2B2} \quad p_{A1A2B2B2} = p_{A1A2} \times p_{B2B2} \quad p_{A2A2B2B2} = p_{A2A2} \times p_{B2B2}$$

This implies that knowledge of an individual's genotype at one locus provides no information on its genotype at the other locus. There are several possible causes for a deviance from independence. Physical linkage among the considered loci is one reason: in this case, recombination among the two loci is not frequent enough to make the two mono-locus genotypes to evolve independently from one another. Clonality and selfing generate linkage disequilibrium throughout the genome of the considered species. Population structure is another origin of linkage disequilibrium among loci, as an extension of the Wahlund effect when considering more than one locus. Finally, selection may create linkage disequilibrium through epistatic effects*.

4.2.3.4 Isolation by Distance

Isolation by distance (IBD) occurs whenever the amount of gene flow exchanged among sub-populations decreases with the geographic distance separating them. Rousset (1997) optimized the detection of IBD while examining what its signal could tell us about demographical processes. He considered matrices of the geographic distances (Gd) separating any two populations of a sample and their genetic distance (measured as $F_{ST}/(1 - F_{ST})$), and how this relationship is associated with two demographic traits: D, the average local density of reproducing adults, and σ, the average geographic distance separating the birthplaces of parents and offspring. He showed that under isolation by distance:

$F_{ST}/(1-F_{ST}) \sim a + \ln(Gd)/(4\pi D\sigma^2)$ for all species colonizing two-dimensional habitats, and

$F_{ST}/(1-F_{ST}) \sim a + (Gd)/(4D\sigma^2)$ for species colonizing a linear habitat (such as seashores, rivers or ecotones).

The recent version of the *GENEPOP* software (Rousset 2008) allow users to estimate the slope of the IBD linear regression, giving one direct access to the value of the product $D\sigma^2$.

4.2.4 Detecting Selection

4.2.4.1 Diverse Forms of Selection

The polymorphisms observed at selected loci within and across populations are shaped not only by the joint-action of the evolutionary forces that affect neutral loci (i.e., mutation, migration, genetic drift and mode of reproduction), but also by the precise action of selection. Let us consider a selected locus L with two alleles *A1* and *A2*, where the average number of descendants per genotype differs among the three possible genotypes within the considered populations based on their relative fitness (i.e., ω_{A1A1}, ω_{A1A2}, ω_{A2A2}). Selection is said to be directional when the fitness ranking of genotypes results in an increase or decrease in the frequency of one allele relative to the others (e.g., $\omega_{A1A1} \geq \omega_{A1A2} > \omega_{A2A2}$ selects for allele *A1* and against allele *A2*). This is what happens in the case of 'purifying selection', where each new allele *Ax* appearing by mutation in the considered populations will be selected against relative to the ancestral alleles *A1* and *A2*. The phrase 'positive selection' qualifies the opposite situation, where a new allele *Ax* appearing by mutation in the considered populations will be selected for relative to the *A1* and *A2* ancestral alleles. In this latter case, a 'selective sweep' occurs when the new beneficial allele *Ax* replaces, more or less rapidly, the ancestral alleles.

Diversifying selection corresponds to cases where selection is not directional and thus does not lead to allele fixation. This can be achieved through overdominance, when heterozygote fitness is higher than both homozygote fitnesses ($\omega_{A1A2} > \max[\omega_{A1A1}; \omega_{A2A2}]$), either within all habitats and populations, or on average across distinct habitats. The latter case, which is referred to as marginal overdominance, corresponds to instances where the fitness of the heterozygous genotype *A1A2* falls between the fitnesses of the homozygous genotypes *A1A1* and *A2A2* in all possible habitats and where disruptive selection occurs on homozygous genotypes such that the homozygous genotype selected in one habitat is counter-selected in other habitat(s). As a consequence, the F_{ST} estimates computed for a selected locus that experiences marginal overdominance could be either higher or lower than the F_{ST} estimates obtained for neutral loci regardless of whether or not the habitats involved in disruptive selection have been evenly sampled. Alternatively, diversifying selection can also be achieved through frequency-dependent selection that increases the frequency of rare alleles. Frequency-dependent selection should ultimately stabilize allelic frequencies within sub-populations, hence stabilizing allelic frequency differences (thus F_{ST} estimates at the concerned loci) independently of the joint actions of migration and genetic drift. Frequency-dependent selection regulates allelic frequencies at loci that determine mating types (Milgroom and Cortesi 1999; Uyenoyama 2005) and also at loci such as Major Histocompatibility Complex genes that help determine the efficiency of the vertebrate immune response to parasitic infections (Garrigan and Hedrick 2003).

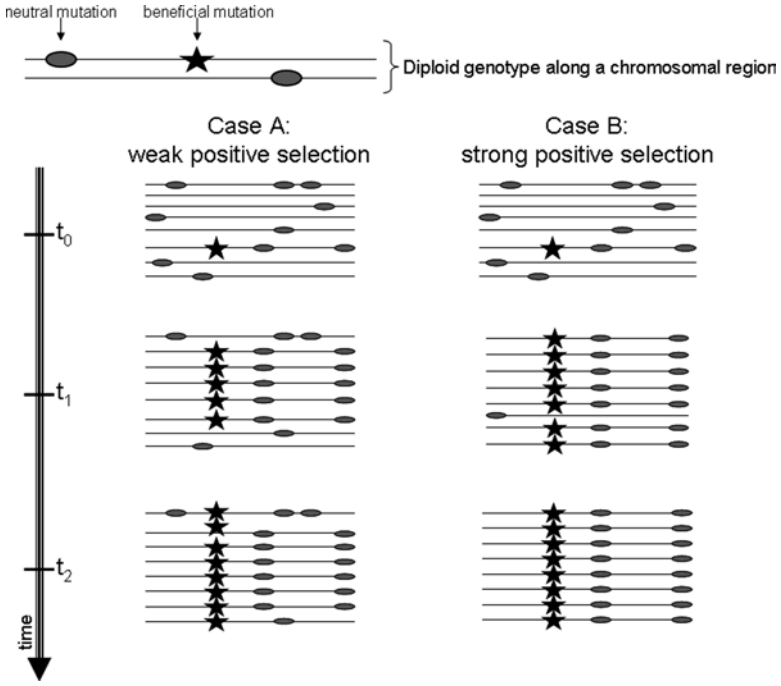

Fig. 4.3 Schematic representation of the effect of positive selection on the polymorphism observed at the locus where a beneficial mutation appeared and at neighbouring loci. *Horizontal bars* represent a small chromosomal region, stars a beneficial mutation that is positively selected once present in the population, and ellipses are neutral mutations. The figure is to be read from *top to bottom* to follow the evolution of within-population polymorphisms over time within the chromosomal region considered. Cases A and B depict two evolutionary trajectories with the same initial neutral polymorphism, but with a higher intensity of positive selection- hence a quicker increase in frequency of the beneficial allele- in B relative to A

4.2.4.2 Signatures of Positive Selection

The apparition of mutations in the genome of parasites conferring resistance to the drugs used to clear parasitic infections are typical examples of positively selected mutations (see below for an example from the human malarial parasite, *Plasmodium falciparum*). Figure 4.3 illustrates the consequences of positive selection for the evolution of polymorphisms in cases of a small (case A) or large (case B) selective advantage conferred by the beneficial mutation that appeared at time t_0. In both cases, the increase in the frequency of the beneficial mutation is associated with a decrease in the neutral polymorphism at loci neighbouring the selected locus on the chromosome. This is the so-called hitch-hiking phenomenon that creates allelic associations which may be broken later, through the effect of recombination or, to a lesser extant, mutation (Andolfato 2001). In Fig. 4.3, recombination between a

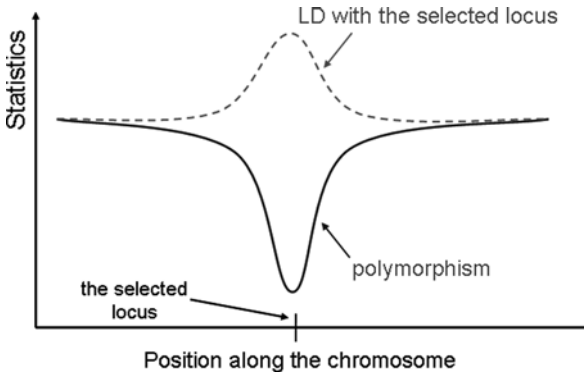

Fig. 4.4 Signatures of positive selection on the chromosomal region where a beneficial mutation appeared (modified from Nielsen 2005)**.** A first signature is associated with the reduction in genetic diversity (i.e., level of polymorphism) observed at a neutral locus when its physical distance to the positively selected locus decreases. A second signature is based on the increase in linkage disequilibrium (*LD*) between the selected locus and a neighbouring neutral locus when the physical distance between them decreases

haplotype bearing the beneficial mutation and another haplotype which does not bear such a mutation only occurs in case A where the population retained a substantial amount of polymorphism early on in the selective sweep. As the probability of recombination between loci increases with the physical distance separating them along the chromosome, the episode of positive selection will ultimately generate two characteristics shown in Fig. 4.4; namely a pattern of reduced polymorphism centred on the selected locus, and a pattern of linkage disequilibrium between the selected locus and neighbouring loci forming a 'bump' that is also centred on the selected locus (Nielsen 2005).

4.2.4.3 A Case-Study: Evolution of Drug-Resistance in *Plasmodium falciparum*

Plasmodium falciparum is the most virulent agent of human malaria. Among the currently available drugs, chloroquine is the one that has been the most intensively and generally used across the distribution area of this parasite, both for protecting humans from new *P. falciparum* infections (prophylaxis) and for clearing existing infections. Cases of chloroquine-resistance were firstly detected at the Thailand/Cambodia border in the late 1950s and in Colombia in the late 1960s (Payne 1987). In the following decade, cases of resistance became frequent all over the African areas where *P. falciparum* circulates (Anderson and Ropper 2005). How many genes are involved in cases of chloroquine-resistance and are they the same across the distributional area of this parasite?

Wootton et al. (2002) addressed this question by genotyping 87 *P. falciparum* parasites from numerous geographic localities (Brazil, Papua New Guinea, African or Asian countries) and with known resistance status (i.e., resistant or susceptible to chloroquine) at 342 microsatellite loci chosen to be evenly distributed among the 14 parasite chromosomes. Wootton et al. (2002) analyzed the variation pattern in genetic diversity among these microsatellite loci (i.e., along the chromosomes) in order to detect the two signatures of positive selection represented in Fig. 4.4. They actually found both signatures when examining microsatellite polymorphisms at loci located nearby a gene located on chromosome 7, the *Pfcrt* gene, that codes for a transporter previously suspected to be involved in chloroquine resistance, hence that had been called PfCRT for *Plasmodium falciparum Chloroquine Resistance Transporter* (Fidock et al. 2000). Wootton et al. (2002) pursued their study by investigating the polymorphism of the *Pfcrt* alleles and the associated encoded proteins. Sequences of other *Plasmodium* species allowed the identification of the ancestral proteins associated with susceptible isolates of *Plasmodium falciparum*. From this point, three mutational events leading to chloroquine–resistance were identified. The first was only identified on parasite isolates from Papua New Guinea and was suspected to have appeared in the mid-seventies. The second resistance allele appeared in 1978 in Brazil. The third appeared in Asia in 1977 and has very quickly spread, by migration, across the Asian and African continents, being almost fixed in these geographic areas by the early eighties. Clearly, coupled to the positive selection defined by drug usage, both independent mutational events and gene flow among populations (even geographically distant ones) explain the dynamics of the resistance to anti-malarial drugs.

4.3 Inferring Transmission and Demographic Functioning of Parasite Populations

Let us imagine that a veterinarian discovers that a steer belonging to a cattle herd is infected by a parasite that has evolved resistance to the drug typically used for curing such an infection. The first question he/she will have in mind concerns the potential for these drug-resistant parasites to be quickly transmitted to the rest of the herd. In other words, the veterinarian will wonder if the unit of reproduction of the parasite is the host population (i.e., the cattle herd) or the steer itself (Fig. 4.5). Differing between the two alternative patterns of genetic transmission is possible when parasite sampling is done at the level of the individual host. For parasites whose unit of reproduction corresponds to the host population (top-panel of Fig. 4.5), Weir and Cockerham's (1984) unbiased estimate f of F_{IS} will have the same average whether it's computed within individual hosts or among host individuals within the population (i.e., absence of Wahlund effect among individual hosts), and the θ estimate of F_{ST} computed among host individuals belonging to the same host population will tend to be null. In contrast, for parasites whose unit of reproduction corresponds to the individual host (bottom-panel of Fig. 4.5), Weir and Cockerham's

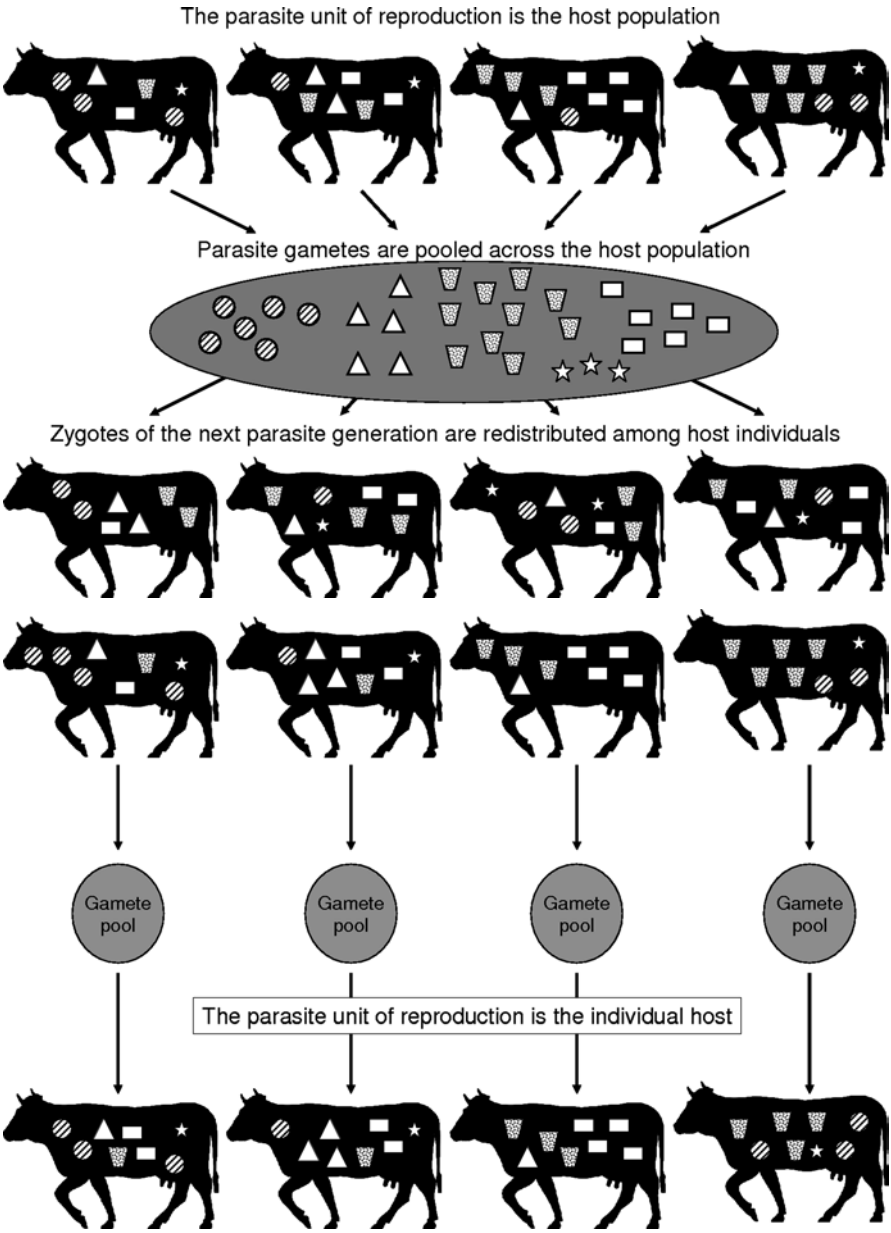

Fig. 4.5 Transmission of parasites' genes between two successive parasite generations. Let us consider parasites infesting cattle. On the *top-panel*, all the parasite gametes produced by the cattle meet and produce the next parasite generation; here the reproductive unit of the parasite is the host population. On the *bottom-panel*, the only parasite gametes that are produced by parasites infesting the same host individual meet; here, the reproductive unit of the parasite is the host individual. As a result, if we ignore the scale of the host individual when sampling parasites from the *bottom-panel*, we would detect Wahlund effect within samples (i.e., a heterozygote deficit that is induced by an inaccurate sampling design)

(1984) unbiased estimate f of F_{IS} computed over the host population will be higher than that computed within host individuals, and the θ estimate of F_{ST} computed among host individuals will be significantly positive as parasite populations infecting distinct host individuals will have diverged due to drift. Study designs that include a detailed sampling of each individual host are thus crucial for studying the population structure of parasites (de Meeûs et al. 2009; Prugnolle and de Meeûs 2010).

Deciphering whether it is the individual host or the host population that defines the parasite population is also crucial for evaluating the degree to which genetic diversity can be maintained within a population of hosts (i.e., for determining the parasite's 'effective population size' or N_E). This is a key parameter because it controls genetic drift (i.e., the higher N_E, the weaker genetic drift) and the potential response to selection (i.e., the higher N_E, the quicker selection operates because of overcoming the action of genetic drift). Criscione and Blouin (2005) formalized the changes in N_E estimates for parasites for the two transmission patterns described in Fig. 4.5, as well as those induced by variation in reproductive success among parasite adults. Everything else being equal, clumped parasitic transmission, where the host individual delimitates the parasite population, inflates the effective population size of the parasite to $N_E = N_C / (1 - F_{ST})$, where N_C is the census size of the focal population (e.g., Rousset 2004, p 14).

Two alternative methods can be used to identify whether parasite populations are delimitated by host individuals or host populations (providing of course that more than one parasite has been sampled per individual host!). As mentioned above, the first relies on changes in Weir and Cockerham's (1984) estimates f and θ when changing the defined scale of a parasite population from the host individual to the host population. The second relies on an extension of Weir and Cockerham's (1984) analysis of variance to three levels (instead of two) structuring the total parasite population, so that the diploid individual parasite represents the smallest level of organization, the individual host the intermediate level, and the host population the largest level for examining the population structure of the parasite. Goudet (2005) recently developed a R package that is precisely designed for testing the occurrence of population structure at three levels of organisation. We demonstrate the use of both methods in the following example of the southern cattle tick.

4.3.1 The Southern Cattle Tick Rhipicephalus microplus in New Caledonia

The southern cattle tick *Rhipicephalus* (*Boophilus*) *microplus* was initially a parasite of bovid hosts (Osterkamp et al. 1999) in India and Indonesia (Labruna et al. 2009), before becoming a major invasive pest in tropical and subtropical agrosystems (Frisch 1999). This is partly due to the recurrent introduction of European breeds of *Bos taurus* in the tropical belt and the quasi-inability of these cattle to mount efficient immune responses against this ectoparasite. Therefore, control of *R. microplus* populations in tropical *B. taurus* herds has largely been achieved

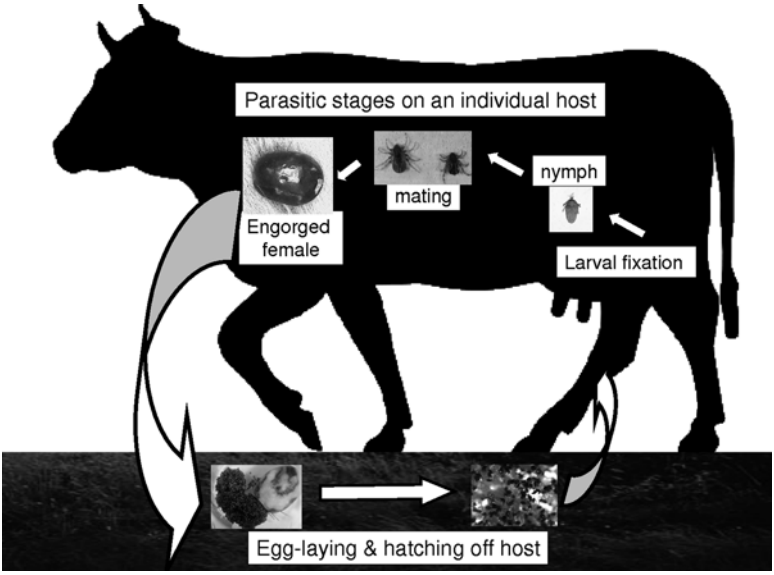

Fig. 4.6 Theoretical life cycle of the southern cattle tick, *Rhipicephalus (Boophilus) microplus*. Most of the life cycle of this ectoparasite occurs on the host. Fully engorged females drop off the host to releasing eggs on the ground and die. Eggs hatch and young larvae survive up to 15 days while seeking a new host. The parasitic part of the life cycle involves larvae, nymphs and adults; three stages that are all blood-feeding with the exception of adult males. *R. microplus* can complete three to four generations per year in New Caledonia

through intensive acaricide use on cattle. As a consequence, *R. microplus* ticks have recurrently developed acaricide resistance in most acaricide-controlled populations (Frisch 1999). The recurrent evolution of acaricide resistance has also been observed at local scale, such as on the Pacific island of New Caledonia.

New Caledonia was free of large herbivores until the mid-nineteenth century when European colonists developed a cattle production industry using British and French breeds of *Bos taurus*. A century after their arrival on the island, *R. microplus* was accidentally introduced with importation of Australian livestock, and started to quickly invade all cattle herds within the island (Verges 1944; Rageau and Vergent 1959). This ectoparasite is generally said to use a single and unique host individual during the parasitic part of life cycle (i.e., Fig. 4.6). Under this assumption, the most likely limits for one *R. microplus* population should be the individual host (i.e., steers and cows within cattle herds).

For characterising the demographical functioning of this invasive parasite in New Caledonia, *R. microplus* was collected according to a hierarchical sampling design involving from three to five individual steers per herd for a total eight cattle herds spread over the island (Fig. 4.7). The pattern of polymorphism for six microsatellite loci within and among four nested levels (individual tick, ticks sampled on one individual host, cattle herd and total) was analysed by Koffi et al. (2006).

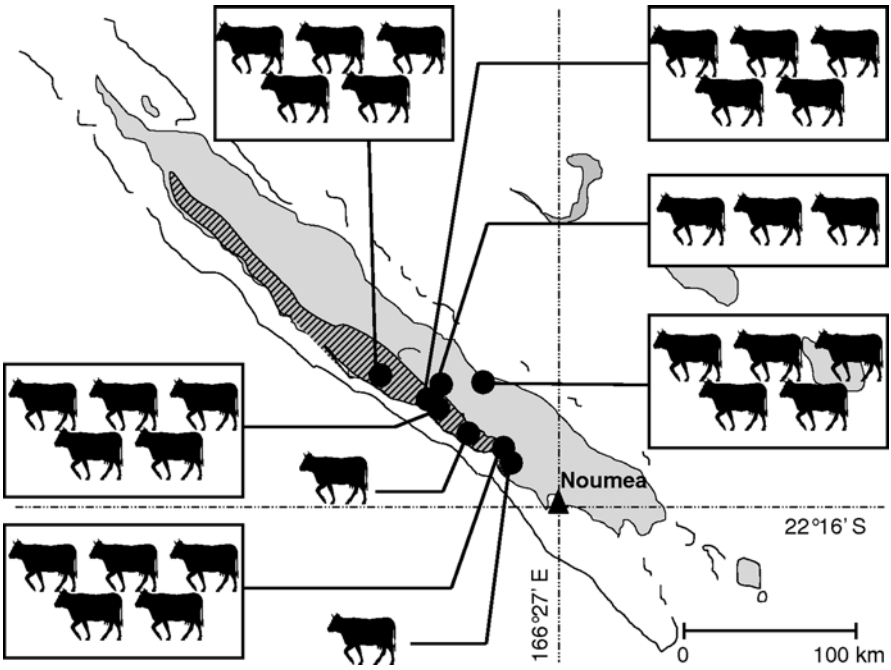

Fig. 4.7 Hierarchical sampling design of *R. microplus* in New Caledonia. The hatched area on the western coast represents a plain where cattle production is the most intensive. The figure details the tick samples analyzed in Koffi et al. (2006). Ticks were sampled on N = 1 to N = 5 individual hosts within each of eight cattle herds, with variation in tick sample sizes per individual host of between N = 4 and N = 24

No clear evidence of population structure within cattle herds was detected; Weir and Cockerham's (1984) estimates of F_{IS} assuming that tick populations were delimitated by cattle herds ($0.010 \leq F_{IS\text{ - within - herd}} \leq 0.017$) were not greater than estimates derived by assuming that tick populations were delimitated by individual hosts ($0.030 \leq F_{IS\text{ - within - host}} \leq 0.071$). The re-analysis with HIERFSTAT (Goudet 2005) confirmed the absence of a Wahlund effect (or substructure) within herds, that is, tick populations were delimitated at the by herd level. The following estimates of structure were found at each level of organization of the tick populations:

- the identity between the two alleles present within ticks relative to that of alleles from two ticks exploiting the same individual host was estimated by $F_{tick\text{-}cows}$ with $0.030 \leq F_{tick\text{-}cow} \leq 0.071$;
- the identity among alleles sampled from two ticks exploiting the same individual host relative to that of ticks exploiting two steers of the same herd was estimated by $F_{cow\text{-}herd} = 0.001$, which was not significantly different from zero ($P > 0.05$);
- the identity by descent among alleles from two ticks exploiting two steers of the same herd relative to that of ticks exploiting steers from distinct herds was estimated by $F_{herd\text{-}total} = 0.014$, which was significantly different from zero ($P < 0.01$).

Therefore, contrary to the usual claims of veterinarian textbooks, the tick *R. microplus* tends to change of individual hosts within cattle herds during its life cycle (movements that are probably promoted by the grooming behaviour of cattle at night).

Koffi et al. (2006) completed their analysis by two complementary investigations aimed at characterizing the demographical functioning of the southern cattle tick. First, they detected a significant signal of isolation by distance corresponding to a weak increase in $F_{ST}/(1–F_{ST})$ with the logarithm of geographical distance (km) between tick samples (slope $b = 0.00054$). As the slope b is also equal to $[4\pi D\sigma^2]^{-1}$ with D representing the density of reproducing adults per tick population and σ the average distance separating the birthplace of parents to offspring (see Sect. 2.3.3), estimating b enables one to calculate the product $D\sigma^2 = [4\pi b^2]^{-2} \sim 147.36$. From then, independent information on either D or σ^2 is required for getting access to estimates of both parameters. For *R. microplus* in New Caledonia, the carrying capacity per steer is about 100 adult ticks per steer while the average density of steers per herd is about 80 steer/km² in the island. This means that the upper-bound for D is locally about $D_{max} \sim 8{,}000$ adult ticks per square kilometre. Given the constraint on the product $D\sigma^2$ resulting from the signal of isolation by distance, such a upper-bound on D defines a lower bound on σ given by $\sigma_{min\text{-}due\text{-}to\text{-}Dmax} = 0.136$ km. Likewise, cattle herds typically rotate among two to five pastures in New Caledonia, with a mean total grazing surface of about 2.5 km². As a consequence, two successive tick generations are likely to be borne on different pastures, and thus more than $\sqrt{0.5}$ km apart. In other words, $\sigma_{min\text{-}due\text{-}to\text{-}farming\text{-}practices} = 0.707$ km constitutes a lower bound for σ in New Caledonia. Given the constraint on the product $D\sigma^2$, such a lower-bound on σ defines a upper bound on D given by $D_{max\text{-}due\text{-}to\text{-}farming\text{-}practices} \sim 295$ adult ticks/km². All together, given the $D\sigma^2$ estimate and the extreme values of D and σ^2, the demographical functioning of *R. microplus* populations in New Caledonia looks characterised by an adult tick density of about a few 100 reproducing ticks per square kilometre and by a gene flow distance per tick generation about 1 km.

Second, Koffi et al. (2006) also tested whether *R. microplus* populations on this island have retained some signature of the recent colonization event from a likely small sample of Australian ticks (~244 tick generations before the population genetics study). This founding effect corresponds to a genetic bottleneck; i.e., a temporary but drastic reduction in population size. The number of alleles present in a bottlenecked population remains lower, even when it has recovered is normal census size, than expected on a demographical stable population (i.e., under mutation/drift equilibrium) of the same census size. As a consequence, until a new mutation/drift equilibrium is reached, the genetic diversity index measured in the bottlenecked population (i.e., the expected frequency of heterozygotes H_E) will be above H_{EQ}, with H_{EQ} representing the genetic diversity index computed for a population with the same number of alleles but at mutation/drift equilibrium (Cornuet and Luikart 1996). Regarding *R. microplus* in New Caledonia, bottleneck signatures were actually detected at each of the six microsatellite locus assayed and within all the tick populations analysed (Koffi et al. 2006), suggesting that the quarantine, which has

been set for any animals entering into New Caledonia since 1945, has efficiently protected the island from multiple re-introductions of *R. microplus*. Cornuet and Luikart (1996) showed that, with less than ten loci and given the *Ne,* the local effective population size after the bottleneck, their method can only detect bottlenecks that occurred between $\tau = (2Ne)/10$ and $\tau = 2Ne$ generations before the analysis. Given that *R. microplus* completes an average of four generations per year in New Caledonia, if the detected bottleneck corresponded to the introduction of the tick in New Caledonia in 1942, then the effective size of the first naturalized tick population would be $120 \leq Ne \leq 1200$. Given that a tick populations seems correspond to a cattle herd and that a cattle herd grazes in average on 2.5 km² in New Caledonia, this *Ne* range in first naturalized tick populations is interestingly similar to the maximal density of reproducing ticks that we previously estimated, for the actual tick populations, by analyzing the local farming practices and the signal of isolation by distance ($D_{\text{max-due-to-farming-practices}} \sim 295$ adult ticks/km², thus a maximal number of reproducing tick per herd of about 738 adults). Alternatively, if the detected bottleneck resulted from current pesticide controls of *R. microplus*, the recovering tick populations would have much lower effective sizes (for instance $1 \leq Ne \leq 10$ for $\tau = 4$ generations $\equiv 1$ year), leading to high inbreeding within the pesticide controlled populations. This is interestingly not the case given the low F_{IS} estimates found for ticks within cattle herds ($0.010 \leq F_{\text{IS-within-herd}} \leq 0.017$).

Interestingly, indirect evidence that the New Caledonian populations of *R. microplus* were large and connected to each other by little gene flow was found by a survey of the physiological diversity in resistance to pyrethrinoids and amitraz acaricides; i.e., the two most commonly used chemicals in tick-control along the two last decades (Chevillon et al. 2007b). Indeed, while we would have expected a resistant mutant to have quickly invaded all tick populations of the island if tick populations were small and frequently exchanging genes, this survey detected different resistant genotypes for each of these chemicals in different New Caledonian tick populations suggesting that resistance evolved several times in different island localities.

4.3.2 Ixodes uriae, a Circumpolar Tick Exploiting Colonial Seabirds

Ixodes uriae has an atypical distribution that covers sub-polar regions in both the Arctic and the Antarctic, where it parasitizes birds from very divergent colonial seabird species (Guiguen 1988). Its lifecycle resembles that of *R. microplus* (Fig. 4.7) in that it involves three blood-feeding life stages (larvae, nymphs and adult females). However, *I. uriae* can use different host individuals for each blood meal and requires from 2 to 4 years to complete its life cycle as it can typically only blood-feed once a year when seabirds return to their colonies to reproduce. In this way, this tick spends most of its lifetime in the substrate of seabirds' nests waiting for the next seabird

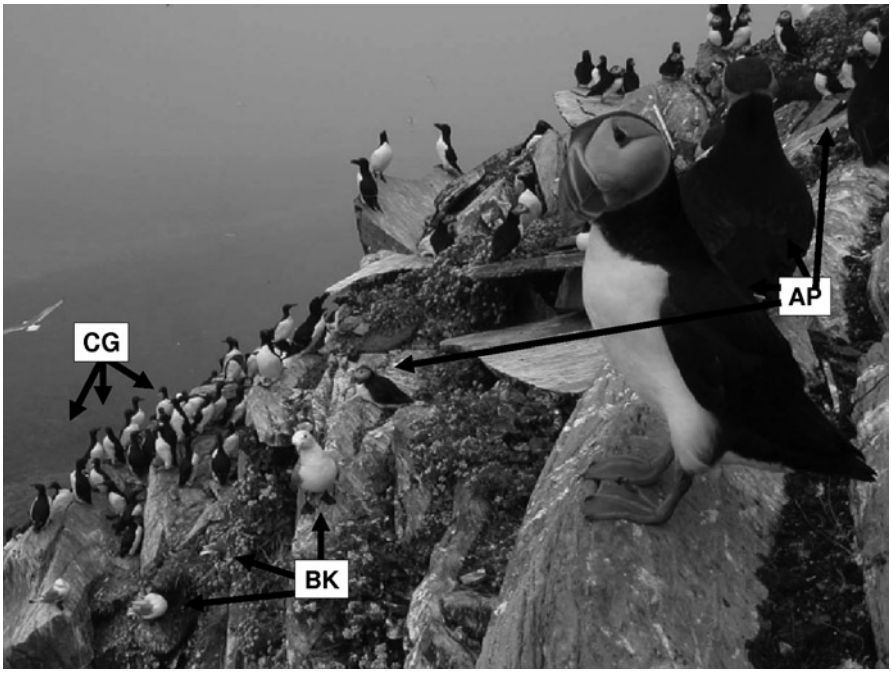

Fig. 4.8 An example of a multi-specific seabird breeding colony on Hornøya, an island in northern Norway. Black-legged kittiwakes (*BK*), Atlantic puffins (*AP*) and Common Guillemots (*CG*) nest in high densities in this colony and despite subtle differences in nesting preferences, can be largely sympatric. *Ixodes uriae* ticks are frequently found exploiting these different species, but have been shown to have formed distinct host-associated races (see text)

breeding season. The population biology of *I. uriae* has been more intensively studied in the northern hemisphere where it is frequently found on seabird species such as the black-legged kittiwake (*Rissa tridactyla*), the common guillemot (*Uriae aalge*) and the Atlantic puffin (*Fratercula arctica*), species which often share large breeding colonies (Fig. 4.8). Given its wide host range, this tick was initially considered as a generalist ectoparasite of seabirds. McCoy and colleagues tested this assumption by analyzing patterns of population structure among ticks sampled from different seabird species breeding in sympatry. To do so, they sampled about 30 ticks per host species per colony and typed these ticks at eight microsatellite loci (Fig. 4.9). Significant heterozygote deficits relative to HWE were observed at almost all combinations of microsatellite loci and seabird colonies when assuming that each colony represented a single tick population irrespective of host use within the colony (average $F_{IS} = 0.099$, $P_{combined} < 0.0001$, $n = 9$ colonies). Such deviations from HWE disappeared when explicitly considering the seabird species on which ticks were collected, suggesting the occurrence of a Wahlund effect within multi-specific seabird colonies (McCoy et al. 2001). This hypothesis was confirmed by a multivariate

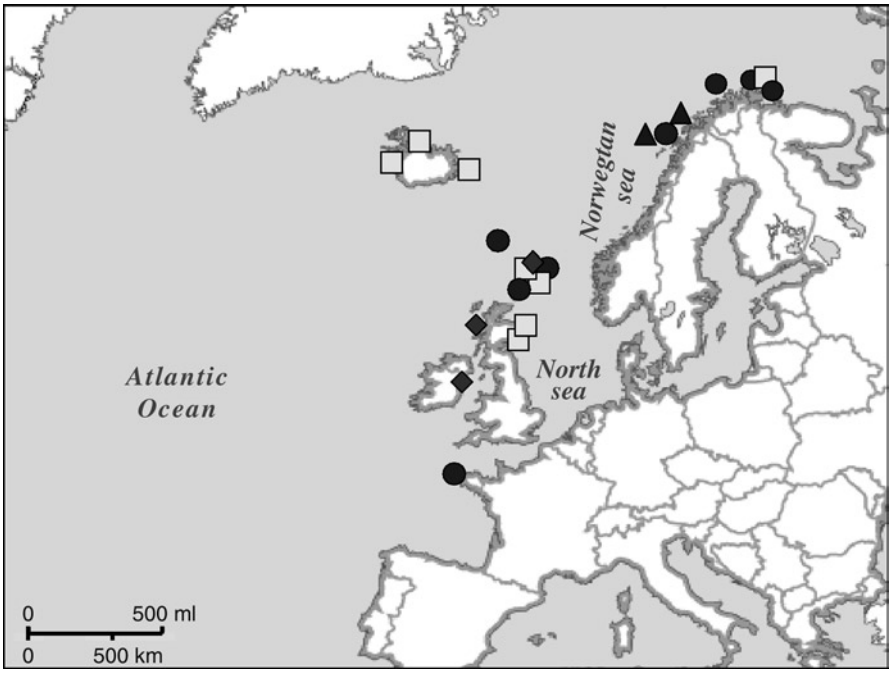

Fig. 4.9 Geographic distribution of *Ixodes uriae* sampling in the sub-Arctic. Thirty ticks per host species present at each breeding colony were sampled. The sampled seabird colonies are represented by *squares* were multi-specific (i.e., with several different seabird species breeding sympatrically) and are represented by *circles, triangles* and *diamonds* when only involving black-legged kittiwakes, Atlantic puffins or common guillemots, respectively

analysis investigating the main pattern of genetic differentiation among tick samples (see left panel in Fig. 4.10). From this point, the demographic functioning of each of the three tick host-races was investigated independently. Different demographic patterns then emerged for each tick host race; an absence of isolation by distance (see Sect. 4.2.3.4) for ticks exploiting Atlantic puffins and significant positive signals for the two other tick host races considered, though with different slopes if all colonies are included (see right panel in Fig. 4.10; McCoy et al. 2003). More recently, these authors analyzed the pattern of genetic diversity among host races and locations using sequences of a more slowly evolving mitochondrial gene in order to test whether host-associated divergence in this tick was old enough to be detected in both mitochondrial and nuclear genomes (Kempf et al. 2009a). Figure 4.11 shows that the divergence of these tick host races has been too recent to see a signature in mitochondrial polymorphism.

In Antarctic regions, *Ixodes uriae* parasitizes numerous different seabird species, including penguins such as the king penguin (*Aptenodytes patagonicus*), the macaroni penguin (*Eudyptes chrysophylus*) and the rockhopper penguin (*E. chrysocome*).

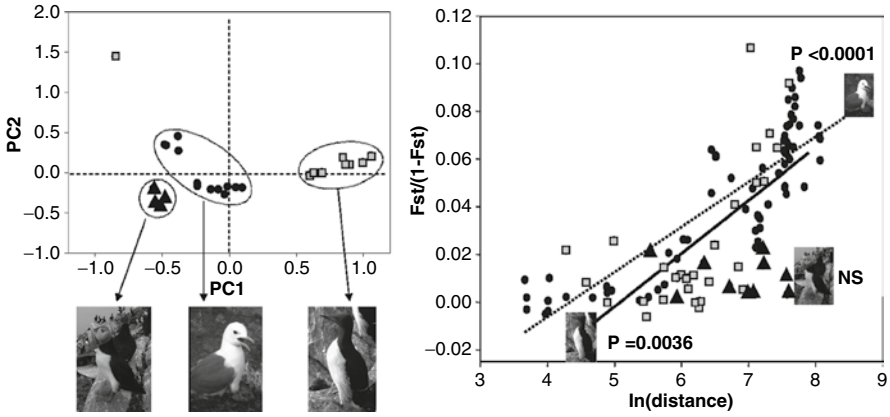

Fig. 4.10 Host-dependent population structure of *I. uriae*. The *left panel* shows, through a principal component analysis (with axes *PC1* and *PC2* representing highly significant parts of the genetic diversity, $P < 0.001$ in both cases), the degree to which the multi-locus genotypic composition of the tick depends on the species identity of its host; there are at least three host races present in the North Atlantic populations of *Ixodes uriae* (Figure derived from McCoy et al. 2005). The *right panel* shows the signal of isolation by distance obtained for each of these three tick host races. The relationship is significant for the ticks exploiting either black-legged kittiwakes or common guillemots, but the slopes differ for the two host-associated groups if all colonies are included (Note: the outliner population of guillemot ticks from the left panel has been removed). In the case of the Atlantic puffin tick host race, there is no evidence of a pattern of isolation by distance (See text or McCoy et al. (2003) for further details)

Interestingly, the sub-Antarctic populations of this tick have also diverged into local host-associated races (McCoy et al. 2005). In other words, host races have evolved at least twice (i.e., at least once per hemisphere) in this tick species!

Ixodes uriae is not only an ectoparasite which imposes deleterious direct effects on its host's fitness (Boulinier and Danchin 1996), but it can also vector a variety of pathogenic agents (Dietrich et al. 2010), including the bacteria responsible for Lyme disease; i.e., the bacterial species included in the species complex *Borrelia burgdorferi sensu lato* (Richter et al. 2006). It is clear that the divergent evolution of ticks into host-specialized races will likely have a significant impact on the transmission of these bacteria. Indeed, (Gomez-Diaz et al. 2010) recently showed significant differences in infection intensity for B. *burgdorferi sl* among three different Arctic host races of this tick, differences which should a significant impact on the transmissibility of the bacteria to new host individuals (Fig. 4.12). To summarize, the tick species *Ixodes uriae*, formerly considered as a generalist prior to population genetics analyses, has been found to be an example of a parasite with hidden host-associated local structure, a structure which may play important role in the transmission risk and global circulation of Lyme borreliosis.

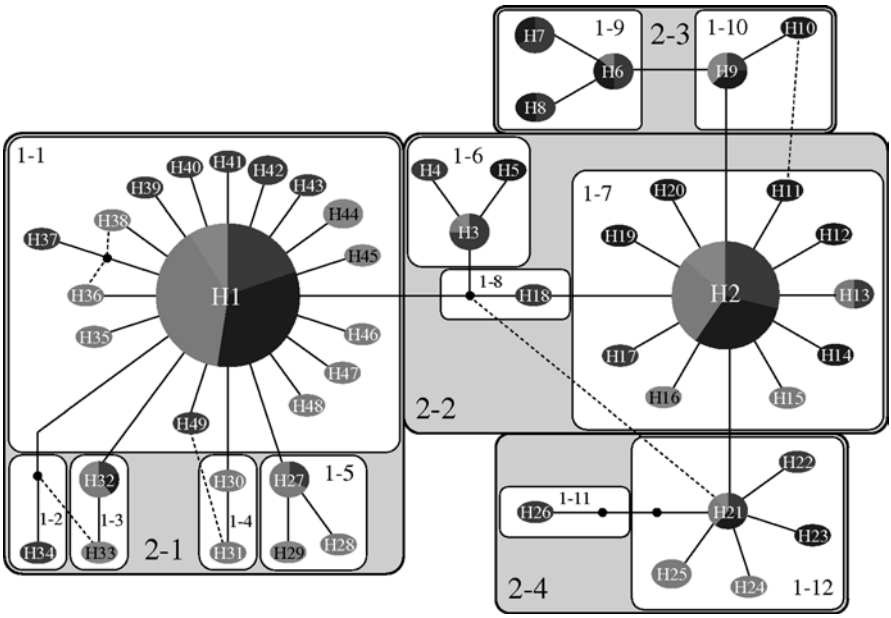

Fig. 4.11 Analysis of *Ixodes uriae* mitochondrial diversity in relation to host race. Mitochondrial sequence haplotypes are designed by alpha-numeric codes (e.g. H1, H44, H46, etc.) corresponding to the individual tick of origin. The four levels of *grey* refer to the seabird host species sampled in the North Atlantic (Altantic puffin, common guillemot, black-legged kittiwake, and razorbill *Alca torda*). Identical sequences are included within *circles*, so that the circle size increases with the frequency of a given sequence. Circles connected by a line correspond to sequences differing by one base pair, while squares delimitate distinct groups (or clades) of related sequences. It is obvious that similar sequences do not group by colour (hence, by host species). In other words, the association between host use and mitochondrial diversity is non significant (contingency test, P = 0.076 see Kempf et al. 2009b for details). Given the slower evolution of mitochondrial sequences relative to microsatellites, the absence of host-associated signals in mitochondrial diversity in contrast to the strong signals obtained with microsatellite markers (Fig. 9) indicates that the divergence into host races within *Ixodes uriae* is rather recent

4.4 Inferring the Reproductive Systems of Parasites

4.4.1 Bisexual Parasites: Testing for Panmixia vs Heterogamy-Homogamy

Sampling couples is possible in parasite species where males guard the females with which they mate. This is, for instance, the case in *Schistosoma mansoni* where females are placed within males' 'gynecophorous canal' (literally the channel carrying the female, see Fig. 4.13). Sampling mating pairs makes it possible to compute the genetic relatedness between the two mating adults and compare it to the

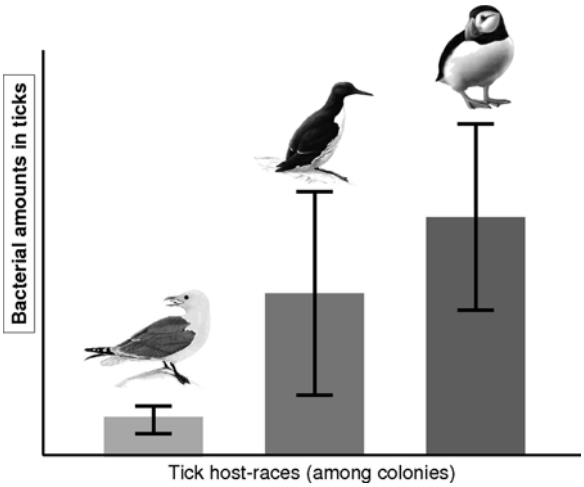

Fig. 4.12 Host race differences in the quantity of bacteria per tick. The relative number of *Borrelia burgdorferi* bacteria per tick was estimated through a quantitative PCR procedure that targets a 145 base-pair zone embedded within a chromosomal gene coding for a conserved structural protein, the flagelline. Average infection intensities did not vary among ticks sampled in different colonies from the same host species, but did vary significantly among host races

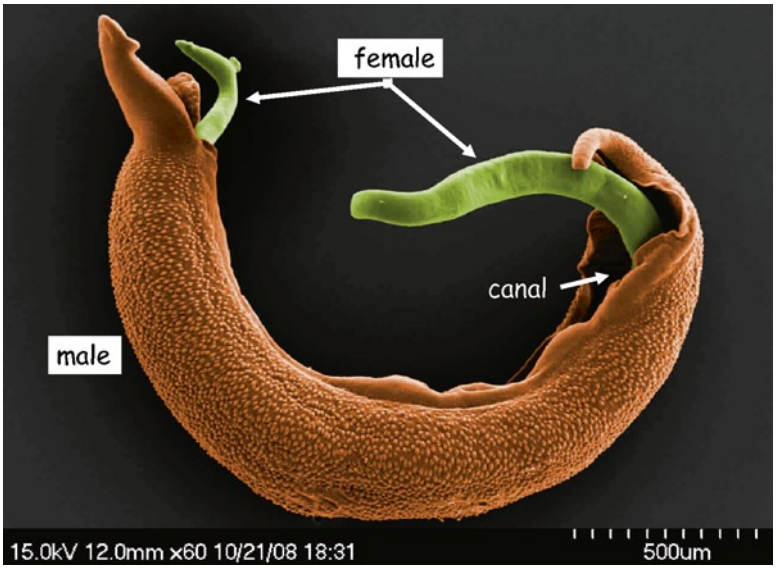

Fig. 4.13 A couple of *Schistosoma mansoni* couple viewed in electronic microscopy. The male is figured in *orange*. The female, figured in *green*, is found embedded within the male's gynecophorous canal. Picture kindly provided by EMTRIX: A Globally Accessible Electron Microscopy Facility, The University of Montana, Missoula, MT 59812, USA

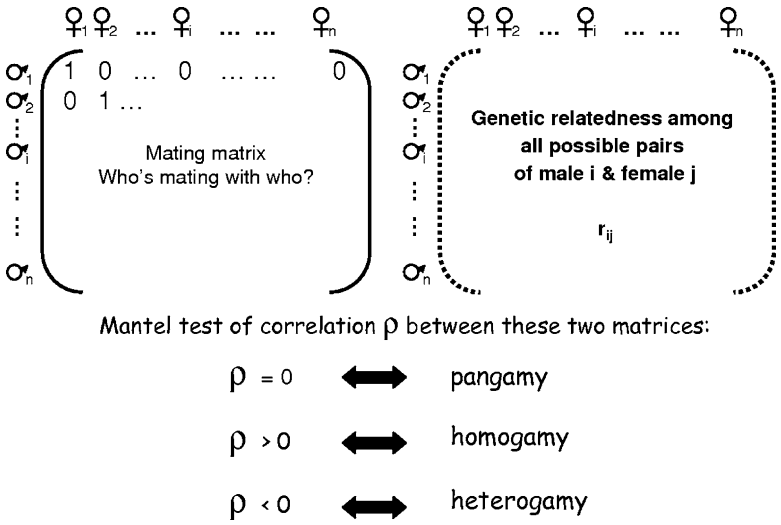

Fig. 4.14 Testing for pangamy when it is possible to sample mating pairs. In both matrices, the lines are defined by males and the columns by females with both genders ranked so that the diagonal corresponds to the mating pairs

distribution of male-female relatedness among all possible mating pairs (or a random sub-sample of these pairs) that one could form within the sampled population. Pangamy corresponds to the case where mating pairs are randomly formed, thus neither more nor less genetically similar than any possible male-female pair from the same population. Pangamy contrasts both with homogamy and heterogamy, where adults respectively tend to mate with the most and the least genetically similar individuals from their population.

When mating adults are sampled within a population and genotyped at the same loci, two matrices can be designed (Fig. 4.14). The first describes which males were mating with which females: $a_{ij}=1$ if the ith male was mating with jth female and $a_{ij}=0$ otherwise. The second contains estimators of pairwise genetic relatedness between all sampled males and all sampled females. From here, a Mantel test can easily be performed for testing whether these two matrices are positively correlated (expected under homogamy), negatively correlated (expected under heterogamy) or uncorrelated (expected under panmixia). This type of analysis has been successfully applied to test for pangamy in populations of *S. mansoni* infesting black rats (*Rattus rattus*) in Guadeloupe (French West Indies) (Prugnolle et al. 2004), *Rhipicephalus microplus* infesting *Bos taurus* cattle in New Caledonia (Chevillon et al. 2007a) where pangamy could not be rejected, and *Ixodes ricinus*, the main vector tick of Lyme disease in Europe where homogamy was shown in half of the studied populations (Kempf et al. 2009b).

4.4.2 An Example of a Parasite with a Complex Life Cycle Involving Clonal and Sexual Reproductive Modes

Plasmodium falciparum is the most virulent agent of human malaria. As all *Plasmodium* parasites, its life cycle is complex (Fig. 4.15) and involves multiple rounds of clonal reproduction for a single round of sexual reproduction, with diploid stages only being found in the mosquito vectors (*Anopheles sp*.).

The population genetic structure of this parasite is often performed on haploid stages collected in human blood. Based on this type of sampling, scientists have usually claimed populations of *P. falciparum* to be 'clonal' (hence far more inbred than expected under HWE) in regions of low malaria transmission, and 'panmictic' in regions of high malaria transmission. However, complementary studies, performed on the diploid stages of malaria agents, were required to confirm the supposed random fusion of gametes. Razakandrainibe and colleagues performed such a study by trapping mosquito vectors in 11 villages in Kenya where malaria transmission remains high throughout the year. They dissected mosquitoes to search for oocysts in the midgut, and typed the extracted oocysts at seven microsatellite loci (Razakandrainibe et al. 2005). Among the 4,602 dissected mosquitoes, 181 carried oocyst stages with an average 6.0 ± 0.7 oocysts per mosquito. In the end, 613 oocysts carried by 145 mosquitoes were successfully genotyped at the microsatellite loci. Significant linkage disequilibrium was detected for each of the 21 possible pairs of these loci ($P < 0.05$); a result congruent with the frequency of clonal reproduction during the parasite's life cycle (Fig. 4.15). F_{ST} and F_{IS} estimates among infected midguts (i.e., the population unit considered here is the mosquito midgut where gametes eventually meet, and where the diploid stages are present) were both significantly positive ($P << 0.01$). The F_{ST} estimate was $\theta \sim 0.36$ (with 95% confidence interval bounded by 0.33 and 0.38) among infected midguts, and the F_{IS} estimate within midguts was $f \sim 0.147$ (95% confidence interval bounded by 0.062 and 0.209). Given the *Plasmodium* life cycle (Fig. 4.15), selfing is the only explanation for such high heterozygous deficits relative to HWE within infected vectors. In this kind of a case, the relationship $s = 2F_{IS} / (1 + F_{IS})$ can give the selfing rate s which is the probability of fusing male and female gametes produced by the same diploid individual (Hartl and Clarke 1989; de Meeûs et al. 2007). Therefore, parasitic individuals are 43% inbred (~0.43 probability of identity by descent), with such high inbreeding being mostly explained by the subdivision of *P. falciparum* into numerous small infra-populations to which 25% of selfing (or something equivalent to selfing) is added. Such results indicate that this parasite is far from panmictic even in regions where it is intensively transmitted, and that it evolves in highly structured populations linked to both infected vectors and human hosts. This structure favours the maintenance of high genetic variability among parasites circulating within the same village. A crude rationale for this observation is that genetic drift operates at the level of infected individuals (mosquitoes and humans), such that allelic loss occurs within, but not among, infected individuals. This rationale has recently been formalized by Prugnolle et al. (2010).

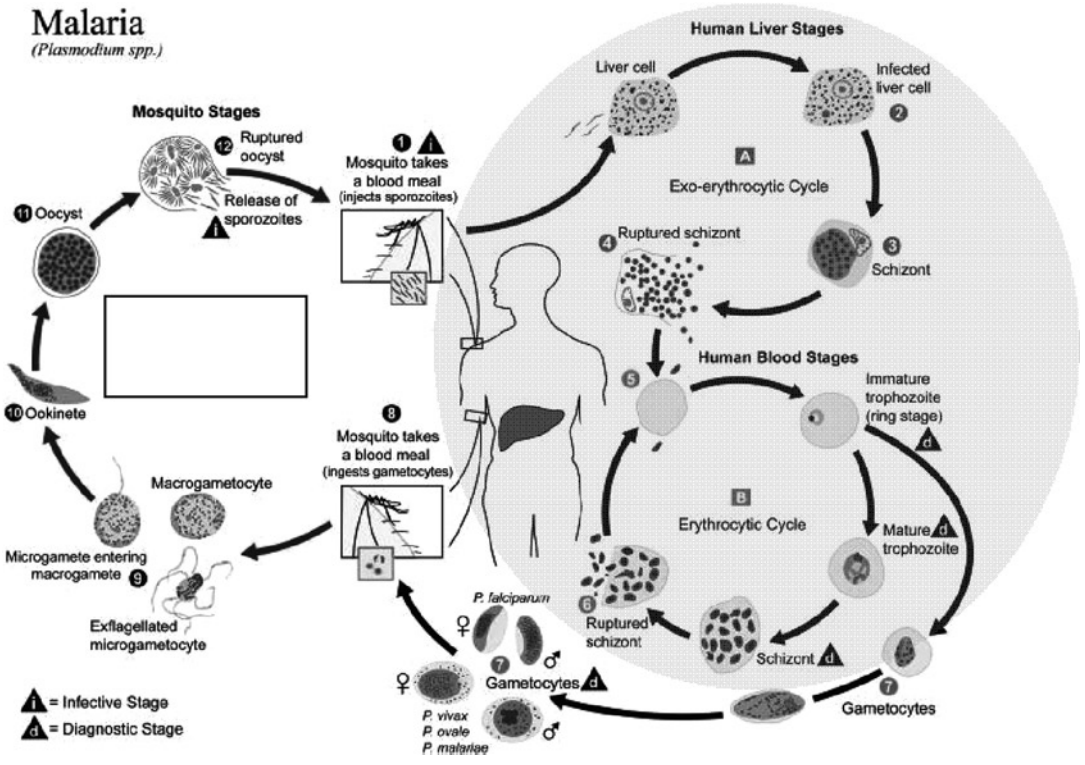

Fig. 4.15 Life cycle of *Plasmodium* parasites. A multiplicity of parasitic stages occur within mosquito (*on left*) and human organs (*on right*). Most arrows connecting successive stages involve clonal reproduction; they appear here below a *gray* shadow. The only step of sexual reproduction is bounded by the cellular differentiation of gametocytes and the formation of the diploid and unicellular zygote referred to as ookinete. This ookinete with the multi-cellular stage called oocyst are the only two steps of the life cycle for which cells are diploid. Ookinete and oocysts are thus the only two stages where scientists have direct access to the result of fecundation (Modified from a scheme retrieved from Wikimedia commons)

4.4.3 *Detecting Clonality as the Main Reproductive Mode Used by a Parasite; an Example of Candida albicans*

Candida albicans is a diploid opportunistic yeast present in the gastrointestinal and genitourinary flora of most healthy humans and other mammals. It is only seriously pathogenic in immuno-compromised patients such as those infected with HIV (Hull et al. 2000; Berman and Sudbery 2002). The sexual cycle was demonstrated under experimental conditions but, how diploidy is restored, how often it occurs in nature, and what kind of population structure this yeast displays, remain largely unknown.

Such questions were recently re-investigated using a sampling design where HIV-infected patients were sampled five times, instead of once as usually performed (Nébavi et al. 2006). This study demonstrated an increase in F_{IS} estimates when pooling samples across distinct patients relative to the F_{IS} estimates found within patients; indicating the occurrence of a Wahlund effect among human hosts. Interestingly, the change in F_{IS} estimates induced by pooling samples among hosts was more pronounced when pooling among female patients ($N = 19$) than among the male patients ($N = 23$) suggesting that women maintain larger yeast populations than men (Nébavi et al. 2006). In addition to these findings, clonality was identified as the main reproductive mode of *C. albicans* in this dataset through the application of recent analytical tools designed for clonal species (de Meeûs and Balloux 2005; de Meeûs et al. 2006).

The main indices of a clonal reproductive strategy rely on very high heterozygous frequencies, hence a negative F_{IS} estimates at all loci (for details see de Meeûs and Balloux 2005; de Meeûs et al. 2006). It is therefore noteworthy that such estimates can only be derived when the average sample size per parasite population is far above one! More generally, a recent review and theoretical analysis demonstrated other biases related to parasitic reproductive modes that result from sampling one parasite per host individual (Prugnolle and de Meeûs 2010).

4.5 Conclusion

Throughout this chapter, we have seen that population genetics may help identify the genetic bases of parasite resistance to drugs (Sect. 4.2.4.3), characterise the spatial limits of a parasite population (Sects. 4.3.1 and 4.3.2), indicate which the parasite populations are most strongly connected via dispersal (Sect. 4.3.2), detect drastic changes in parasite population sizes (Sect. 4.3.1), and forecast the rules under which neutral polymorphisms evolve within parasite populations (Sect. 4.4). Through various different case studies, we hope to have convinced the reader that the accuracy of such inferences mainly depends on two pivotal factors that are the behaviour of the chosen genetic markers and, even more importantly, the sampling design performed.

Glossary

Bottleneck Population bottleneck refers to drastic reduction in population sizes. This induces, at all loci, a decrease in heterozygous frequency and an even more pronounced drop in the allele numbers observed relatively to a demographical stable population with the same census size. In other words the bottlenecked population will remain away from mutation/drift equilibrium for a number of generations following the reduction in census size. The longer last the bottleneck (in term of generation) and the stronger the demographic reduction, the more intense and durable will be the genetic signature of such an event.

Epistatic effects there are epistatic effects on selection among loci A and B whenever the intensities of selection pressures acting on the polymorphism at locus A vary among the different genotypes observed at the locus B and *vice versa*.

Identity by descent two sampled alleles are said identical by descent if they result from an event of DNA replication in any ancestral generation. The alleles identical by descent are obligatory identical in state.

Identity in state two alleles are identical in state whenever the genotyping methods used is unable to discriminate them.

Founding effect the colonization of a new habitat is usually achieved by a few individuals relatively to the standard range of population sizes in the ancestral distribution area of the immigrating species. Therefore the colonization of a new area, habitat or host species for a parasite, usually induces a population bottleneck; such a bottleneck associated to colonization is called founding effect.

Homoplasy there is homoplasy whenever the identity in state and the identity by descent are not synonymous. Reverse mutation and polymorphism with a finite number of distinguishable alleles (i.e. allelic sizes of microsatellite alleles) are frequent causes of homoplasy.

Fixation an allele has reached fixation in a population when it remains the only allele present within a population. The immigration of different alleles into the population and mutation into a different allelic state are the only ways to re-create local polymorphism in a fixed population.

Mutation/drift equilibrium let's consider an isolated population (no immigration into it, no emigration from it) of constant finite size, and a neutral locus at which pangamy is realized so that it tends to be at HWE. At this locus along time, mutation will regularly introduce new alleles while genetic drift will regularly make existing alleles to disappear. The combined action of mutation and genetic drift will make evolving the polymorphism observed at the considered locus to evolve and reach an equilibrium where the number of distinct alleles (hence the heterozygous frequency) will remain constant even if the identity of the alleles present in the population will keep on changing.

Mantel test The Mantel test is a statistical test of correlation between two matrices of same dimensions that is adapted to the case where the elements of any matrix are not independent from one another (such as the matrix of either geographical

distances or that of genetic distances among populations). This test is performed by randomly permuting the rows and columns of one matrix multiple times.

Panmixia random meeting of the gametes produced in a population; theoretically achieved only for species where individuals are self-compatible hermaphrodites.

Pangamy random mating among the sexual partners present in population; theoretically achieved only for hermaphrodite species.

References

Anderson T, Ropper C (2005) The origin and spread of antimalarial drug resistance: lessons for policy markers. Acta Trop 96:269–280

Andolfato P (2001) Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev 11:635–641

Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol Evol 24:21–30

Berman J, Sudbery P (2002) Candida albicans: a molecular revolution built on lessons from budding yeast. Nat Rev Genet 3:918–928

Boulinier T, Danchin E (1996) Population trends in Kittiwake Rissa tridactyla colonies in relation to tick infestation. Ibis 138:326–334

Brookes A (1999) The essence of SNPs. Gene 234:177–186

Chevillon C, Koffi BB, Barré N et al (2007a) Direct and inferences on parasite mating and gene transmission patterns- pangamy in the cattle tick *Rhipicephalus* (*Boophilus*) *microplus*. Infect Genet Evol 7:298–304

Chevillon C, Ducornez S, de Meeûs T et al (2007b) Accumulation of acaricide resistance mechanisms in *Rhipicephalus (Boophilus) microplus* (Acari: Ixodidae) populations from New Caledonian Island. Vet Parasitol 147:276–288

Cornuet J, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2004

Criscione CD, Blouin MS (2005) Effective sizes of macroparasite populations: a conceptual model. Trends Parasitol 21:212–217

Criscione CD, Poulin R, Blouin MS (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. Mol Ecol 14:2247–2257

David P, Perdieu M-A, Pernot A-F et al (1997) Fine-grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. Evolution 51:1318–1322

de Meeûs T, Balloux F (2005) *F-statistics* of clonal diploid structured in numerous demes. Mol Ecol 14:2695–2702

de Meeus T, Durand P, Renaud F (2003) Species concept: what for? Trends Parasitol 19:425–427

de Meeûs T, Lehmann L, Balloux F (2006) Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. Infect Genet Evol 6:163–170

de Meeûs T, McCoy K, Prugnolle F et al (2007) Population genetics and molecular epidemiology or how to "débusquer la bête". Infect Genet Evol 7:308–332

de Meeûs T, Prugnolle F, Agnew P (2009) Asexual reproduction in infectious diseases. In: Schön I, Martens K, van Dijk P (eds) Lost sex: the evolutionary biology of parthenogenesis. Springer, New York, pp 517–533

Dietrich M, Gomez-Diaz E, McCoy KD (2011) Worldwide distribution and diversity of seabird ticks: Implications for the ecology and epidemiology of tick-borne pathogens. Vector-borne and Zoonotic Diseases 11:453–470

Drummond A, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes – application to human mitochondrial-DNA restriction data. Genetics 131:479–491

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform 1:47–50

Fidock D, Nomura T, Talley A et al (2000) Mutations in the P-falciparum digestive vacuole transmembrane protein PfCRT and evidence for their role in chloroquine resistance. Mol Cell 6:861–871

Frisch J (1999) Towards a permanent solution for controlling cattle tick. Int J Parasitol 29:57–71

Garrigan D, Hedrick P (2003) Detecting adaptive molecular polymorphism, lessons from MHC. Am J Hum Gen 73:363–375

Gomez-Diaz E, Doherty PJ, Duneau D, McCoy K (2010) Cryptic vector divergence masks vector-specific patterns of infection: an example form the marine cycle of Lyme borreliosis. Evol Appl 3:391–401

Goudet J (1995) FSTat version 1.2: a computer program to calculate F-statistics. J Hered 86:485–486

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierachical F-statistics. Mol Ecol Notes 5:184–486

Guiguen C (1988) Anthropozoonoses et oiseaux marins: contribution à l'étude des ectoparasites hématophages des espèces nicheuses sur les côtes françaises continentales et insulaires. PhD thesis (in French), Université Aix-Marseille

Hartl D, Clarke A (1989) Principles of population genetics theory. Sinauer, Sunderland

Hedrick P (2005) Genetics of populations, 3rd edn. Jones and Bartlett Publishers, Sudbury

Hull C, Raisner R, Jonhson A (2000) Evidence for mating of the asexual yeast *Candida albicans* in a mammalian host. Science 289:307–310

Kempf F, Boulinier T, de Meeûs T et al (2009a) Recent evolution of host-associated divergence in the seabird tick *Ixode uriae*. Mol Ecol 18:4450–4462

Kempf F, de Meeûs T, Arnathau C et al (2009b) Assortative pairing in Ixodes ricinus L (Acari: Ixodidae), the European vector of Lyme disease. J Med Entomol 46:471–474

Koffi BB, de Meeûs T, Barré N et al (2006) Founder effects, inbreeding and effective sizes in the Southern cattle tick: the effect of transmission dynamics and implications for pest management. Mol Ecol 15:4603–4611

Kunz W (2002) When is a parasite species a parasite species. Trends Parasitol 18:121–124

Labruna M, Naranjo A, Thompson C et al (2009) Allopatric speciation in ticks: genetic and reproductive divergence between geographic strains of *Rhipicephalus (Boophilus) microplus*. BMC Evol Biol 9:46

Martin D, Williamson C, Posada D (2005a) RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21:260–262

Martin D, Posada D, Crandall K et al (2005b) A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res Hum Retrov 21:98–102

McCoy K, Boulinier T, Tirard C et al (2001) Host-specificity of a generalist parasite: genetic evidence of sympatric host-races in the seabird tick *Ixodes uriae*. J Evol Biol 14:395–405

McCoy KD, Boulinier T, Tirard C et al (2003) Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. Evolution 57:288–296

McCoy K, Chapuis E, Tirard C et al (2005) Recurrent evolution of host-specialized races in a globally distributed parasite. P Roy Soc London B 272:2389–2395

Milgroom M, Cortesi P (1999) Analysis of the population structure of the chestnut blight fungus based on vegetative incompatibility genotypes. Proc Natl Acad Sci USA 96:10518–10523

Nébavi F, Ayala F, Renaud F et al (2006) Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients from Abidjan (Côte d'Ivoire). Proc Natl Acad Sci USA 103:3663–3668

Nielsen R (2005) Molecular signatures of natural selection. Annu Rev Genet 39:197–218

Osterkamp J, Wahl U, Schmalfuss G et al (1999) Host-odour recognition in two tick species is coded in a blend of vertebrate volatiles. J Comp Physiol A Neuroethol Sens Neural Behav Physiol 185:59–67

Pasteur N, Pasteur G, Bonhomme F et al (1988) Practical isoenzyme genetics. Ellis Hortwood Ltd, Chichester

Payne D (1987) Spread of chloroquine resistance in *Plasmodium falciparum*. Parasitol Today 3:241–246

Prugnolle F, de Meeûs T (2010) Apparent high recombination rates in clonal parasitic organisms due to inappropriate sampling design. Heredity 104:135–140

Prugnolle F, Theron A, Durand P et al (2004) Test of pangamy by genetic analysis of *Schistosoma mansoni* pairs within its natural murine host in Guadeloupe. J Parasitol 90:507–509

Prugnolle F, Durand P, Renaud F et al (2010) Effective size of the hierarchically structured populations of the agent of malaria: a coalescent-based model. Heredity 104:371–377

Pybus OG, Charleston MA, Gupta S et al (2001) The epidemic behavior of the hepatitis C virus. Science 292:2323–2325

Pybus OG, Drummond AJ, Nakano T et al (2003) The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. Mol Biol Evol 20:381–387

Rageau J, Vergent G (1959) Les tiques (Acariens: Ixodidae) des îles françaises du Pacifique. Bull Soc Pathol Exot 52:819–835

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and oecumenism. J Hered 86:248–249

Razakandrainibe FG, Durand P, Koella JC et al (2005) "Clonal" population structure of the malaria agent *Plasmodium falciparum* in high-infection regions. Proc Natl Acad Sci USA 102:17388–17393

Real LA, Henderson JC, Biek R et al (2005) Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proc Natl Acad Sci USA 102:12107–12111

Richter D, Postic D, Sertour N et al (2006) Delineation of Borrelia burgdorferi sensu lato species by multilocus sequence analysis and confirmation of the delineation of Borrelia spielmanii. Int J Syst Evol Microbiol 56:s873–s881

Rousset F (1997) Geographic structure and estimation of gene flow from F-statistics under isolation by distance. Genetics 145:1219–1228

Rousset F (2004) Genetic structure and selection in subdivided populations. Princeton University Press, Princeton

Rousset F (2008) GENEPOP007: a complete reimplementation of the GENEPOP software for windows and Linux. Mol Ecol Resour 8:103–106

Uyenoyama M (2005) Evolution under tight linkage to mating types. New Phytol 165:63–70

Verges J (1944) Les tiques du bétail. Méthodes d'éradication. Imprimeries Réunies, Nouméa

Weir B, Cockerham C (1984) Estimating *F-statistics* for the analysis of population genetic structure. Evolution 38:1358–1370

Wootton J, Feng X, Ferdig M et al (2002) Genetic diversity and chloroquine selective sweeps in *Plasmodium falciparum*. Nature 418:320–323

Wright S (1965) The interpretation of population structure by F-statistics with special regard to system of mating. Evolution 19:395–420

Part III
Contribution of Population and Evolutionary Epidemiology

Chapter 5
Modelling the Dynamics of Host-Parasite Interactions: Basic Principles

Pauline Ezanno, Elisabeta Vergu, Michel Langlais, and Emmanuelle Gilot-Fromont

Abstract Mathematical modelling is a valuable tool for the analysis of the infectious diseases spread. Dynamical models may help to represent and summarize available knowledge on transmission and disease evolution, to test assumptions and analyse scenarios, and to predict outcomes of the host-pathogen interactions. This chapter aims at introducing basic concepts and methods of epidemiological modelling, in order to provide a starting point for further developments. After positioning modelling in the process of disease investigation, we first present the main principles of model building and analysis, using simple biological and also mathematical systems. We then provide an overview of the methods that can be employed to describe more complex systems. Last, we illustrate how the modelling approach may help for different practical purposes, including evaluation of control strategies. A brief conclusion discusses the challenge of including genetic and molecular variability in epidemiological modelling.

P. Ezanno (✉)
INRA, Oniris, UMR1300 Bio-agression, épidémiologie et analyse de risque en santé animale, BP40706, 44307 Nantes, France
e-mail: pauline.ezanno@oniris-nantes.fr

E. Vergu
INRA, UR341 Mathématiques et Informatique Appliquées, Domaine de Vilvert, 78350 Jouy-en-Josas, France
e-mail: elisabeta.vergu@jouy.inra.fr

M. Langlais
Institut de Mathématiques de Bordeaux, UMR CNRS 5251, Université Bordeaux Segalen, 3 ter Place de la Victoire, 33076 Bordeaux, France
e-mail: michel.langlais@u-bordeaux2.fr

E. Gilot-Fromont
Université de Lyon, Université Lyon 1, UMR 5558, Bâtiment Mendel, 43 Bd du 11 novembre 1918, 69622 Villeurbanne Cedex, France

Université de Lyon, VetAgro sup, Campus vétérinaire de Lyon, 1, avenue Bourgelat, 69280 Marcy l'Etoile, France
e-mail: e.gilot@vetagro-sup.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_5,

Keywords Deterministic models • Stochastic models • Basic reproductive ratio – R0 • Metapopulation • Disease control • Epidemic models • Structured populations

5.1 Introduction

Daniel Bernoulli was the first to use a mathematical model approach for assessing the effectiveness of prophylaxis method to control the spread of an epidemic disease (Bernoulli 1766, targeted to smallpox propagation). He simplified a rather generic first model based on the fate of a cohort of individuals, keeping only what was clearly needed for his own purposes. Then, Bernoulli performed a mathematical analysis of the resulting model using a celebrated method to solve in closed form a specific class of nonlinear ordinary differential equations (sometimes referred to as Bernoulli's equations). He carried out a sensitivity analysis and evaluated the robustness of his conclusions with respect to the simplifications made and thus tested their relevance (Valleron 2000).

After Bernoulli's pioneer researches, developments in epidemiological modelling mainly occurred in the early twentieth century. Hamer (1906) was interested in the recurrence of measles epidemics. He introduced one of the fundamental ideas in epidemiology, that is, the epidemic spread depends on the rate of contact between susceptible and infected individuals. Hamer formalized this idea using the 'mass action principle', which states that the transmission rate of an infection is proportional to the product of the densities in both susceptible and infected individuals (Anderson and May 1991). At about the same period, Ross (1908) found a relationship between malaria and mosquito abundance. A few years later, the first complete formulation of a generic epidemiological model was proposed by Kermack and McKendrick (1927). The analysis of their model led to the statement of the threshold theorem: after the introduction of a few infected individuals into a fully susceptible population, an epidemic will occur provided the number of susceptible individuals exceeds a critical threshold. Finally, in 1931 Greenwood introduced the idea that chance may intervene in the process of transmission: during a given contact, transmission may occur or not with a certain probability. These three fundamental concepts, contact rate, threshold theorem and randomness in transmission, are at the origin of the modern theoretical epidemiology.

The very first analysis by Bernoulli which encompassed both observations and theoretical hypotheses to predict the effect of vaccination already showed how such an approach is useful and complementary to experimentations and observations. In order to formulate a relevant model, biologists and modellers have to work together to establish the simplest set of rules that summarizes the biological system of interest, according to the objective of the study. The biological side of model building consists in providing observed or experimental data, expert opinion, or knowledge on similar systems, while the modeller perspective is dedicated to choosing and adapting available methods or develop new ones that are appropriate to the system and questions under study. Interaction between biology and modelling is particularly

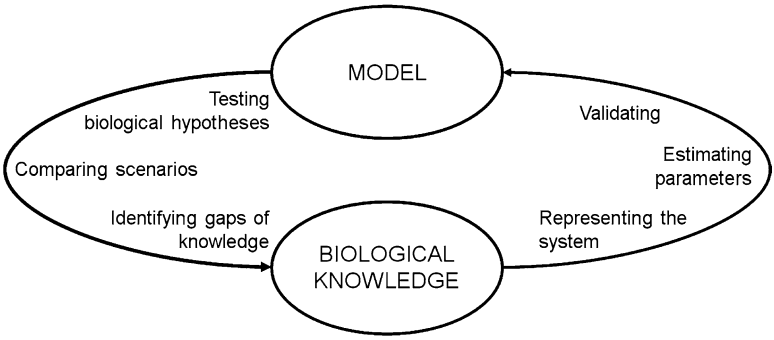

Fig. 5.1 Mutual input of biology and modelling

important to define the scale of the representation (within or between hosts, populations, etc.) and to identify established knowledge together with unlikely or controversial aspects. Hypotheses to be tested can then be formulated. This process gives rise to a mutual input: biology provides the necessary pieces to formalise and validate models, while modelling leads to the formulation of new hypotheses and to the identification of key points and gaps of knowledge that should be investigated through experimental or observational studies (Fig. 5.1).

According to what is known about the system under study and depending on the question to be answered, modelling approaches are used in a variety of manners (Becker 1979; Hethcote 2000; Valleron 2000). The first aim can consist in summarizing available knowledge and constructing a formal representation of the system, in order to facilitate the understanding of underlying complex processes and to provide general qualitative conclusions. Analytical formulations (deterministic dynamical systems, stochastic processes…), computer-based models, graphical schemes or diagrams (conceptual models) are some of possible representations. A classical example of such a descriptive model is the representation of the spatio-temporal dynamics of rabies in wild-living populations (Fig. 5.2): the paradigm of spatio-temporal waves constitutes a reasonable representation of the complex processes underlying rabies expansion, thus it was much used to summarize and explain these processes, including in communication towards non-specialists.

A second aim is to assess the relative importance (essential, secondary or irrelevant) of each of the various mechanisms involved in the system dynamics. With such an objective, an accurate description of the system with clearly stated assumptions and biologically relevant parameters is necessary. Then the model may be used to test biological hypotheses (such as those justifying the structure, parameter values, or the form of transition functions of the model), by comparing the dynamical behaviour of different sub-models including the hypothesis or not. As an example, in order to test whether immune protection or age-dependent infection rates are important processes in the transmission of *Theileria equi* and *Babesia caballi* among horses, Rüegg et al. (2008) compared the goodness-of-fit of predictions issued from

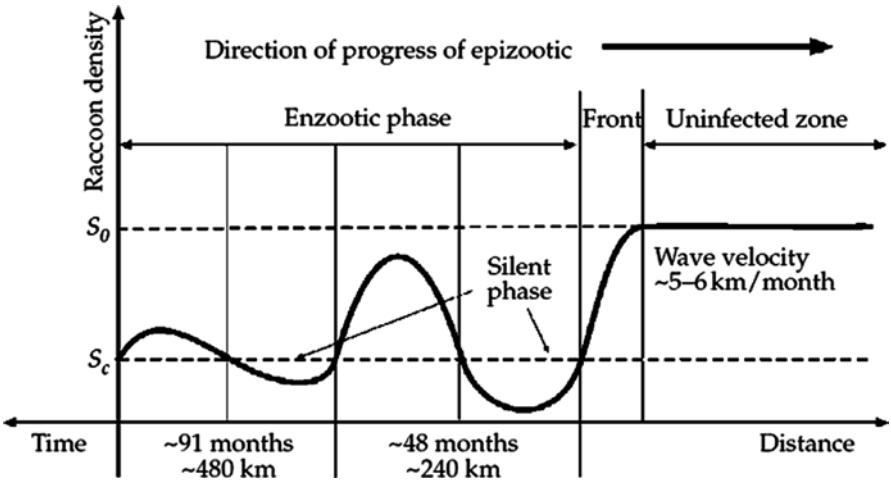

Fig. 5.2 Description of racoon rabies spatio-temporal dynamics through the evolution of host density. Time increases to the *left* of the epidemic front, while distance increases to the *right* of the front (From Real and Childs 2006)

different models. They showed that both mechanisms are important in the transmission dynamics of *B. caballi* (*i.e.*, their inclusion significantly improved the quality of model prediction), but not for *T. equi*.

The same kind of analytical models may also be used to compare different scenarios, and thus answer 'what if' questions. For instance, infection dynamics in different host populations (with different structures, sizes, etc.), in different regions, at different periods, for different variants of a pathogen, etc. can be compared (Becker 1979). Such comparisons cannot generally be carried out in the field or in the laboratory, whereas numerous numerical experiments can be performed by simulation. As an example, Ezanno et al. (2008) evaluated the influence of herd structure on the spread of bovine viral diarrhoea virus within a dairy herd: enhancing contacts between young animals before breeding or isolating lactating cows from other groups both decreased virus spread compared to a herd with a typical structure with indirect contacts between groups.

A last context to use analytical models is the estimation of key parameters from data. For example, the spatial variation of the infection probability for foot-and-mouth disease has been estimated by modelling (Gerbier et al. 2002). Control points of the system and factors of uncertainty that may decrease our confidence in estimations can be identified. Models can also help guiding further data collection in order to improve estimation accuracy.

Third, if the model has been validated against data, it can be used to predict future states of the system depending on observed past ones and on assumptions on mechanisms acting in the future. Whereas quantitative predictions are still subject to some uncertainty even after model validation, qualitative forecasts can be provided for a

variety of situations. For instance the assessment of the relative effectiveness of interventions used to control infectious diseases spread may help to design optimal strategies. This approach was largely used for a variety of diseases such as the foot-and-mouth disease (Ferguson et al. 2001), avian influenza (Boender et al. 2007; Le Menach et al. 2006) and human pandemic influenza (Ferguson et al. 2006; Flahault et al. 2009; Kernéis et al. 2008; Longini et al. 2005 amongst many others).

5.2 Principles of Model Formulation for a Single Homogeneous Population

The formulation and analysis of epidemiological models include several steps: definition of the model structure involving a preliminary choice of a formal representation, analysis of model properties and outputs and identification of thresholds which determine radical changes in model dynamics depending on whether they are exceeded. If an analytical representation of the system in question is chosen, a mathematical formalism should be specified with respect to the context and the question motivating the study. Deterministic or stochastic models in discrete or continuous time are possible representations. Both approaches have their strengths: deterministic models are appropriate in large populations where fluctuations have relatively little overall impact whereas stochastic formulations are more suited for small populations and rare events where randomness has large effects. Although most theoretical aspects presented in this chapter are valid for both categories of models, examples are often related to deterministic models that are more easily described in short terms and appropriate for an introductory text. Stochastic modelling is by no means less relevant for the study of diseases spread and we invite the reader to refer to excellent monographs specifically developing this methodology (Andersson and Britton 2000; Daley and Gani 1999; Keeling and Rohani 2008 to quote only a few). A brief overview of simple prototypes of epidemiological models illustrated by examples is provided in this section, where all the basic models assume a single homogeneous population.

5.2.1 Model Structure

Most epidemiological models start from the description of the infection dynamics at the individual level (Fig. 5.3) to infer pathogen spread at the population level.

At least two individual states should be defined with respect to the disease: susceptible (S) and infected (I). The implicit assumptions when limiting to a two state model are a negligible latency period and an instantaneous return to the susceptible state after infection (this corresponds to the SIS model; Fig. 5.4) or life-lasting infection (the SI model).

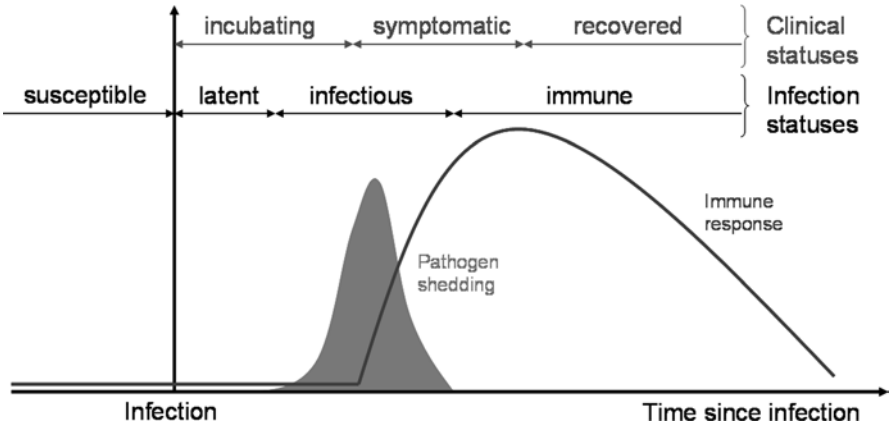

Fig. 5.3 Definition of the individual infection status *vs*. clinical status during a much simplified infection process (Modified from Keeling and Rohani 2008). It has to be noted that the symptomatic period is not necessary simultaneous to the infection status

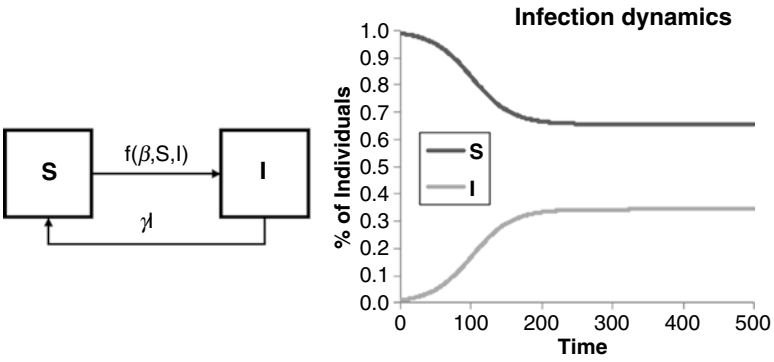

Fig. 5.4 Diagram of a SIS model (susceptible – infected – susceptible) (*left*) and example of simulated dynamics (*right*). The transition from S to I corresponds to the infection process and depends on the transmission rate β, and on the numbers of S and I individuals, f being the force of infection. The transition from I to S corresponds to a recovery (without immunisation) and depends on the mean infection duration ($1/\gamma$ if the sojourn time in compartment I is exponentially distributed)

In order to describe the flows of individuals between compartments within the simplest mathematical framework, additional assumptions have to be made: homogeneous population (*i.e*., in a SI model, all S individuals are of identical susceptibility, all I individuals have identical infectiousness), homogeneous contacts between individuals, a short infection period relatively to the host life expectancy (which allows to neglect population vital dynamics), and a transition from I to S that does not depend on the time since infection. Based on these assumptions, a model may

be formulated in discrete time (*e.g.* the Reed-Frost first models, Daley and Gani 1999; Thrusfield 1995) or continuous time (*e.g.* through a system of ordinary differential equations). When one or several of the above assumptions are not valid, the model should be modified. Typically, if two compartments are not sufficient to describe the infection process, other states may be defined. A recovered (R) status indicates a non negligible delay between the end of the infectious period and the moment individuals lose their immunity (SIRS model). An exposed status (E) is to be considered when there is a non negligible delay between infection and infectiousness. Classically, the probability of recovery per unit of time is constant, which corresponds to an exponential distributed infectious duration (a classical assumption routinely used for mathematical tractability-based reasons). However, more realistic probability distributions can be used. For instance, a gamma distribution can be assumed for the infectious duration by replacing the single previous I compartment by a series of n stages, where the rate of transition between stages is equal to $n\gamma$ (the mean duration of the infectious period being still equal to $1/\gamma$, according to the known property stating that the sum of n equally exponentially distributed variables of mean $1/n\gamma$ is a gamma distributed variable of mean $1/\gamma$).

5.2.2 Model Analysis

The qualitative behaviour of the analyzed system may first be deduced from the mathematical properties of the model. For both discrete and continuous time system models, the resulting mathematical analysis fits into the framework of dynamical systems, a well known and developed branch of mathematics. An important preliminary question is the non-negativity of solution components that is also part of the validation of the rationale underlying model building. Forward invariance is a convenient tool to answer this question.

A second step is the understanding of the population dynamics behaviour before the parasite introduction. Various paradigms are available: constant population size, logistic or mono-stable behaviour (regulation toward a limited carrying capacity), Allee effect or bi-stable behaviour (existence of a population size threshold separating population extinction *vs.* regulation), as well as exponential growth or time-periodic dynamics, to name a few. The exploration of population dynamics requires sorting stationary states (constant solutions) of the model, that can be achieved either in closed form or by using suitable software for numerical or algebraic computation.

Then, a local stability analysis (LAS) should be developed. This consists in assessing whether a small initial departure from a given stationary state will result into the model driving back the population to the original stationary state, or else, driving it toward a new stationary state or to some new horizon, *e.g.* periodic *vs.* chaotic dynamics. A generic mathematical methodology goes through devising a dedicated matrix made of partial derivatives of the model system – referred to as Jacobian matrix of the system evaluated at the given stationary state – and then

computing its spectrum, that is its real and complex eigenvalues. The system is locally stable when all eigenvalues are negative or have negative real parts. Global stability analysis (GAS), that is stability for any nonnegative departure from a given stationary state, is much more complicated since it requires more sophisticated mathematical tools such as building Lyapunov functional. Although complex eigenvalues can make non-mathematician modellers feel uneasy, they are a nice tool to support the existence of oscillatory behaviours in transient solutions as well as to exhibit periodic solutions when time gets large (*e.g*. Hopf biburcation).

The main step now arises: what are the likely outcomes after introducing an infective individual into a naive population? From a mathematical point of view, most questions and answers are identical to those in the foregoing paragraph: stationary states, stability analysis, transient and long time behaviour, put in a somewhat different setting. A first question is whether infection will persist after the introduction of a few infective individuals, *i.e*., LAS of the stationary state without any latent, infectious and immune individuals, the so-called disease free equilibrium (DFE). Computing eigenvalues of a suitable Jacobian matrix will yield an answer (see above). Mathematically this will select a (nonlinear) combination of the parameter set from the model that is to be compared to 0, negativity implying LAS while positivity yields instability (see also R_0 in the next subsection). Heuristically, one may expect the emergence of a LAS endemic stationary state (with infectious individuals) as soon as the DFE becomes unstable, due to one or several parameter(s) variation. This step requires looking for all possible LAS endemic states. What is mostly expected is a *forward bifurcation*, that is the emergence of a unique LAS endemic stationary state when the DFE looses its stability. Dynamics can be much more complex with a *backward bifurcation*, that is the existence of an extraneous LAS endemic state right before the DFE looses its stability (see also the R_0 limit in the next subsection). This means that two different dynamics and LAS regime can coexist: a DFE one and an endemic one, which could make the control of the disease quite uneasy. It may also happen that a LAS endemic state may loose its stability yielding oscillations and time periodic dynamics (cf. rabies model). Numerical simulations can be supported by a suitable mathematical analysis (*e.g*. Hopf bifurcation). For exponentially growing populations (before introduction of the disease) no DFE can exist but the actual question is whether the disease can control and regulate the given population. For time periodic population dynamics (before introduction of the disease), the existence and stability of both DFE and periodic endemic states is challenging to prove from a mathematical point of view, though sophisticated theoretical and numerical tools are available (Bacaër and Gernaoui 2006).

Besides mathematical analysis, numerical simulations are conveniently used to observe or guess transient and long-time dynamics, especially for models using a large number of parameters and state variables. As an example, studying host-macroparasite systems, Rosà et al. (2003) showed that, while a deterministic system predicts oscillatory behaviour, taking into account stochastic events in the system dynamics leads to larger oscillations that may threaten parasite or host persistence (Fig. 5.5). Interactions between model components can be studied, which may give rise to unexpected behaviours. Simulations first require adequate parameterisation of the model, using demographic or epidemiological data (Becker 1989). Then a

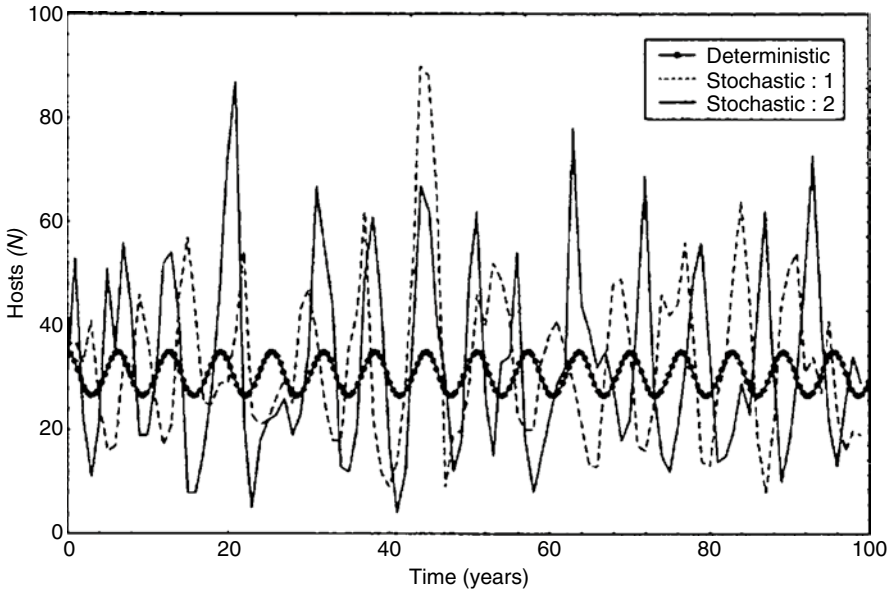

Fig. 5.5 Deterministic prediction and two stochastic simulations for the evolution of the number of hosts across time in the red grouse – *Trichostrongylus tenuis* model (From Rosà et al. 2003)

sensitivity analysis is a relevant approach to evaluate how model outputs vary according to variations in model inputs (parameter values, functions, model structure, etc.; see Saltelli et al. 2000 for a review of methods of sensitivity analysis). First, such an analysis is useful to test modelling assumptions: what if other functions had been chosen? What if the model were simpler with less state variables? etc. A sensitivity analysis also constitutes the first step before using a model to evaluate strategies of control of the system. Only parameters that significantly influence model output and can be managed on the field are potential control points of the modelled system (Ezanno et al. 2007).

5.2.3 *Reproductive Numbers: R_0 and Related Threshold Parameters*

The basic reproductive number, R_0, is one of the most important concepts in epidemiology, population dynamics and ecology provided by the mathematical thinking. Generally speaking, R_0 is the expected number of secondary individuals generated by a typical individual during its lifetime. The term "secondary" depends on context: it means "secondary cases" in epidemiology (where a typical individual refers to an infectious one) and "offspring" in ecology and demography (Heffernan et al. 2005; see Heesterbeek 2002 for a historical perspective on R_0).

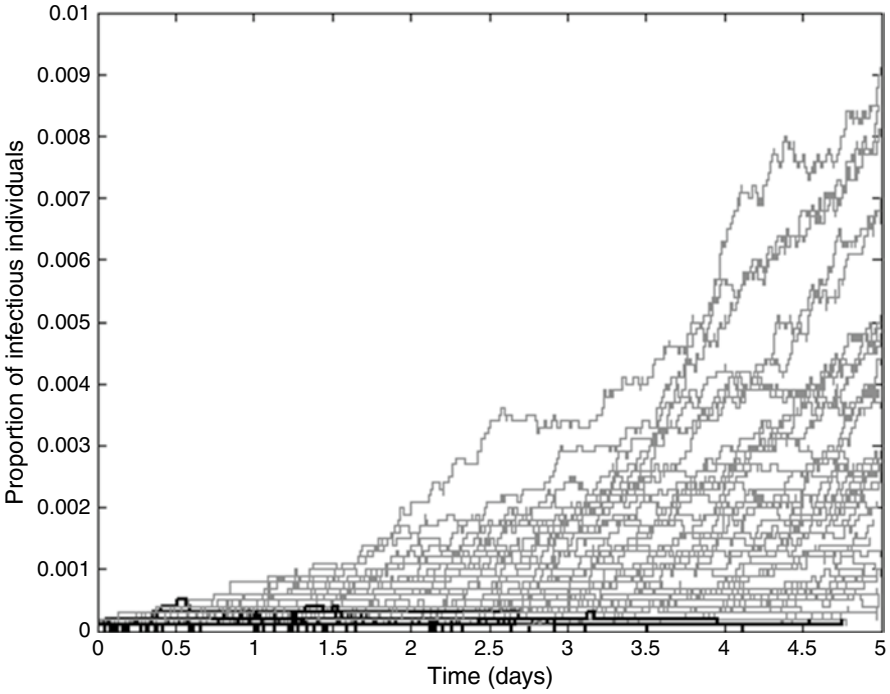

Fig. 5.6 R_0 and epidemic dynamics: proportion of infectious hosts for 100 simulations of an epidemic starting in a susceptible population of size 1,000 after the introduction of one infectious, using a SIR model with $R_0 = 2$. *Black lines* represent simulations where extinction occurs before major epidemics (defined as epidemics leading to infections of more than 90% of the initial susceptibles)

The threshold behaviour of R_0 renders this parameter very useful for predicting the emergence of a new epidemic, the outcome of a started outbreak and the efficacy of mitigation strategies. In a deterministic formulation, if $R_0 < 1$, a pathogen introduced in a completely susceptible population will not be able to invade, whereas if $R_0 > 1$ an epidemic can occur. In a stochastic framework, the interpretation of R_0 is less straightforward: a value of R_0 below 1 predicts the extinction with probability 1, whereas if $R_0 > 1$, invasion is not the only possible outcome because the probability of extinction is not equal to 0. More specifically, in the case of an entirely susceptible population where a single individual is initially introduced, and the dynamics is described by a branching process (with three state variables S, I and R), the probability of extinction before a major epidemic is equal to $1/R_0$ for $R_0 > 1$ (Fig. 5.6). Beyond the context of an entirely susceptible and homogeneous population, other formulations of the reproductive number exist, as briefly stated below.

A general method, developed by Diekmann et al. (1990) (and then described in detail in Diekmann and Heesterbeek 2000, illustrated on a number of specific models in van den Driessche and Watmough 2002 and clearly summarized in Heffernan et al. 2005) allows deriving R_0 as a function of model parameters for a variety of

situations where the population is split into disjoint classes. This method is based on the next generation matrix and R_0 is then calculated as its dominant eigenvalue. Situations as diverse as the following ones are handled by this approach (see van den Driessche and Watmough 2002 for more details): heterogeneous populations divided in several groups with specific behaviours with respect to the disease, such as sexually transmitted diseases; multi-strain systems, such as influenza spread, where several strains may co-circulate and are in competition for the same hosts; vector-borne diseases where both vectors and hosts dynamics should be considered simultaneously. For any of these types of models, a next generation matrix can be built after having identified and separated terms defining new infections. More precisely, for each compartment i, F_i is defined as the rate of appearance of new infections and V_i as the difference between out and in transfers of individuals by any other means. The next generation matrix is then equal to FV^{-1}, where F and V are the matrices of partial derivatives of F_i and V_i respectively, under some particular conditions met by F_i and V_i. Applied to a simple SIR model with no demography, this method would yield to $R_0 = \beta/\gamma$ (see Fig. 5.4 for parameter definition).

The R_0 value derived in this way and more exactly its position with respect to 1 helps assessing questions such as the capability of a pathogen introduced into a fully susceptible population to generate an epidemic. As stated in the previous subsection, for dynamical systems, this can also be assessed by studying the LAS of the DFE which reduces to finding a relationship between parameters that makes all the Jacobian matrix eigenvalues to be negative or have negative real parts. It is important to note that, while these two approaches are qualitatively equivalent with respect to the answer concerning the invasion of the host population by the pathogen, the latter one may supply a threshold parameter different from R_0 (Roberts 2007). Roberts and Heesterbeek (2003) emphasize that if this threshold does not have the same biological interpretation as the dominant eigenvalue of the next generation matrix, it "can therefore not be called the basic reproduction ratio nor denoted by R_0". It should also be pointed out that both methods can define algebraic threshold parameters with no sound epidemiological interpretation.

If the population is not entirely susceptible, the appropriate term to be calculated is the effective reproduction number, R_{eff}, which is equal for an SIR system to $\beta S/\gamma N$, where S is the size of the susceptible population and N is the total population size. More generally, while R_0 is uniquely defined for a couple pathogen/host population, R_{eff} may change over time, as the proportion of susceptible hosts S/N varies.

Beyond derivations of R_0 and R_{eff} as combinations of model parameters in order to identify threshold criteria, various methods were also developed for estimating the value of R_0 and its related variants from data. One of the simplest methods, valid for SIR models and closed populations, connects R_0 to the final epidemic size: $R_0 = \ln\left(s\left(\infty\right)\right)/\left(s\left(\infty\right)-1\right)$ where $s(\infty)$ represents the final proportion of susceptible hosts (Diekmann and Heesterbeek 2000). This assumes that the epidemic is observed until the end which is not always the case. Moreover, it is also of great importance not only to provide a posterior characterization of the epidemic, but mostly to assess the epidemic intensity at its very beginning, in order to tailor mitigation strategies. This is possible for instance by calculating R_0 from r, the initial rate of the exponential

growth of the number of infectious individuals. Several expressions relate R_0 to r, depending on the distribution of the generation time, W (defined as the delay between the moment one individual becomes infected and he/she infects another individual; Roberts and Heesterbeek 2007; Svensson 2007; Wallinga and Lipsitch 2007; Yan 2008). For a simple SIR model and assuming a constant infectivity, $R_0 = 1 + r * T_W = 1 + r * T_I$, where T_W and T_I are the mean generation time and the mean duration of the infectious period respectively. R_0 is certainly of great interest, but it is often more convenient to focus on R_{eff}, since it reflects the actual capability of the epidemic to progress over time and provides information on the impact of control measures in real time. The estimation of R_{eff} reduces to a simple counting of secondary cases if all infected individuals are traced until their index case ("who infected whom" chain). However, most often, this information is not available. In this case, R_{eff} can be estimated for instance by fitting a transmission model to data (Riley et al. 2003). A statistical approach that avoids mechanistic assumptions was proposed by inferring "who infected whom" from the observed curve and times of symptoms by using pairs of cases instead of the entire infection network (Wallinga and Teunis 2004, estimations for the 2003 SARS epidemic in several geographic locations). The scenario where not all secondary cases have been detected was tackled through a Bayesian approach by Cauchemez et al. (2006) who estimated the reproduction number in an ongoing epidemic for the 2003 SARS epidemic in Hong Kong.

As already stated, information on R_0 and its related variants provides valuable insights mainly in two situations: for the evaluation of the invasion risk of a host population by a pathogen and for evaluating and comparing control strategies. For both cases, the choice of an appropriate method for estimating R_0 or R_{eff} should be done with respect to data and objectives and comparison of estimations to previous values and interpretation of discrepancies (if any) should also be provided.

Despite its incontestable role in handling infectious diseases spread, R_0 could sometimes be mis- or overused. Roberts (2007) draws our attention on some exceptions where the basic statements generally fulfilled by R_0 and cited in this subsection are not true (see also the backward bifurcation in the previous subsection). The author cites several mechanisms allowing persistence of an endemic infection even for $R_0 < 1$ (for example assuming that exposure to infection accelerates the transition from the exposed to the infectious state) and points out the existence of situations where the evolution of a pathogen does not necessarily maximise its R_0. Another important point raised by Roberts (2007) and initially fully described in Roberts and Heesterbeek (2003) and then in Heesterbeek and Roberts (2007) concerns structured populations where interventions are targeted at specific subpopulations. In this case, R_0 is less useful and should be replaced by T, the type-reproduction number. Both R_0 and T exhibit the threshold behaviour and are equivalent in homogeneous population, but T is more appropriate in heterogeneous populations since it summarizes the control effort required to eliminate an infection when measures are applied to a specific host type (rather than to the entire population). Finally, care has to be taken when evaluating control efficacy through R_0 values: as pointed out by

Heffernan et al. (2005), since sometimes the use of R_0 could ignore other issues such as the potential negative effect of interventions on population, it is important to simultaneously consider other indicators (the total morbidity or mortality) in addition to R_0.

5.3 More Realistic Models for Complex Situations

The simple models cited in the previous section give a very general idea of two processes involved in epidemiological dynamics: transmission and immunity. However, these may be not sufficient if pathogen dynamics is affected by other traits of the host population, such as heterogeneity among individuals, demographic processes or population structure. Here, we provide a brief overview of further possible models for more complex systems.

5.3.1 Heterogeneity Among Individuals

Individuals in a population do not equally contribute to infection dynamics. First, infectious individuals do not equally shed the pathogen, either because shedding routes are numerous and possibly not simultaneous, or because excretion depends on infection duration or other individual characteristics such as age, genetics, physiological stage, etc. Second, susceptible individuals do not have equal susceptibility, due to individual intrinsic characteristics or previous exposure to the pathogen. Compared to naïve (never exposed) individuals, those that have already been exposed to the pathogen often have a reduced susceptibility and infection duration. Depending on the reduction in susceptibility after a first infection, such a model goes from an SIS model (no protection) to an SIR model (full protection). The same approach is also adequate to represent immunity acquired by vaccination (Glass and Grenfell 2003; Greenhalgh et al. 2000). Lastly, a cross protection may arise when many variants of a given pathogen co-circulate in the population for example (Restif and Grenfell 2007). These heterogeneities in infectiousness and susceptibility interfere with pathogen spread and control both in non-structured (Lloyd-Smith et al. 2005b; Matthews et al. 2006) and structured (Ball and Lyne 2001) populations. Therefore, heterogeneity should be considered, especially when specific individuals are targeted by a surveillance or control program.

Such heterogeneities can be modelled through several methods. A first way is to consider as many categories of S or I individuals as necessary to describe variability in susceptibility (Fig. 5.7) or infectiousness. This has been used to model the spread of human tuberculosis considering that individual may have either a susceptible or a resistant phenotype, assumed to be consistent for an individual over time (Murphy et al. 2003). Another way is to use partial derivative equations when susceptibility or infectiousness continuously varies among individuals (Novozhilov 2008; Veliov 2005).

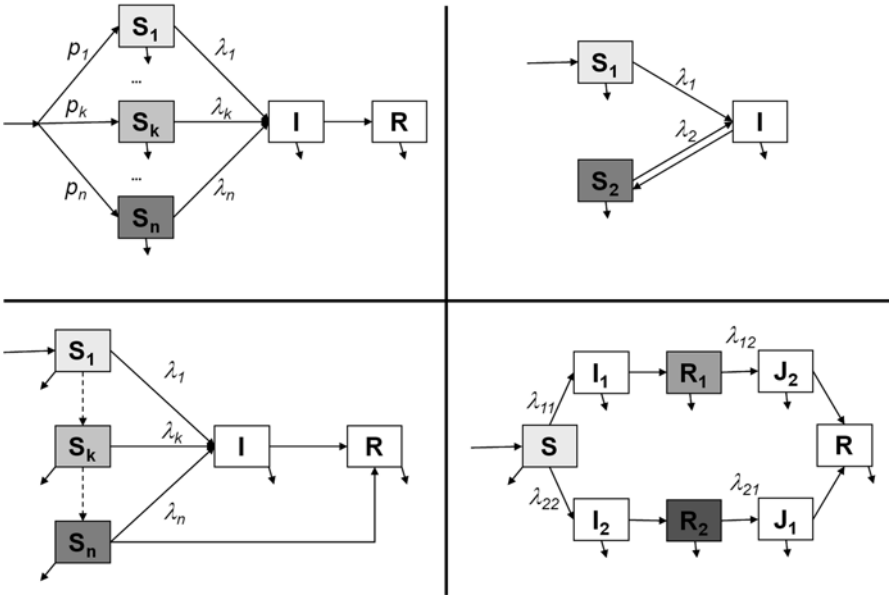

Fig. 5.7 Diagrams of SIR models accounting for heterogeneity in susceptibility. *Top left*: the level of susceptibility varies among individuals but is consistent over life (a proportion p_k of individuals are born with susceptibility k and thus has a force of infection λ_k); *bottom left*: susceptibility variation over lifetime; *top right*: reduced susceptibility following exposure to infection ($\lambda_2 < \lambda_1$); *bottom right*: reduced susceptibility because of cross-protection after infection with a close variant of the pathogen

5.3.2 *Accounting for Population Dynamics*

The assumption that demographic processes may be neglected in front of epidemiological dynamics cannot hold in all situations, specifically when the duration of infection is of the same order of magnitude as life expectancy, but also when considering the long-term dynamics of the system, instead of a single epidemic process. In this case, birth, death and migratory processes should be included in the model. This may be done by adding input and output flows to each compartment.

Including demographic processes in epidemiological models may first help to investigate their influence on disease dynamics. Generally speaking, birth acts to replenish the pool of susceptible individuals and thus to favour disease spread, while mortality has the opposite effect. This has been studied in plant diseases: when crop growth is taken into account, it first entails a dilution effect on leaf lesions, followed by an increase in R_0 due to higher density in susceptible host tissue (Ferrandino 2008). The situation becomes more complex when density-dependent processes occur, which is the case in most natural-living populations: then fecundity and mortality are not independent from population sizes. This gives rise to complex effects, including the possibility that the threshold theorem is not longer valid or observable (Lloyd-Smith et al. 2005a). Disease may also be a major determinant in host

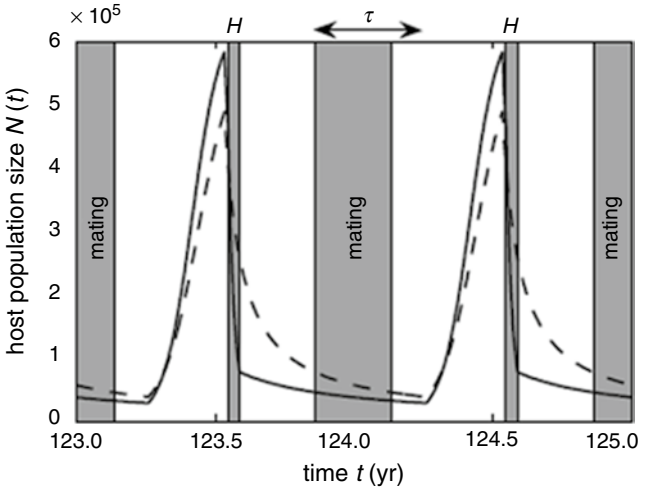

Fig. 5.8 Effect of harvesting on the total host population size in wild boar infected by Classical Swine Fever (From Choisy and Rohani 2006). Numerical solution of the model for a 2-year period, after stationary dynamics has been reached. The periods of mating and harvest (*H*) are represented in *grey*, and the duration of gestation is represented by the length of the *double arrow* above the graph. The *dashed curve* represents the total host population size in the absence of harvesting and the *full line curve* represents the dynamics of the total host population size in the presence of harvesting

population dynamics through its effect on fecundity or survival. HIV infection is one of the leading examples when population demographics need to be accounted for, the virus being able to turn population growth rates from positive to negative values (Anderson et al. 1988). Another case when host population dynamics should be considered is represented by animal populations managed by humans. Rapid changes in density, demographic parameters, spatial distribution and contact structure may result from management decisions and affect disease transmission. Complex dynamics may arise from the interactions between demographic and disease processes. A recent example is the study of hunting on transmission of classical swine fever in wild boar *Sus scrofa* (Choisy and Rohani 2006, Fig. 5.8). The model showed that the drastic reduction in population density due to harvesting results into an overcompensation due to density-dependent birth and death rates. After the next birth period, the population reaches high density, which results in a high level of disease transmission and prevalence. Overall, harvesting is predicted to increase disease spread.

5.3.3 Pathogen Spread in Structured Populations

Beside their heterogeneity, individuals are clustered in groups within which preferential contacts occur. Age, social groups, households, schools or herds strongly

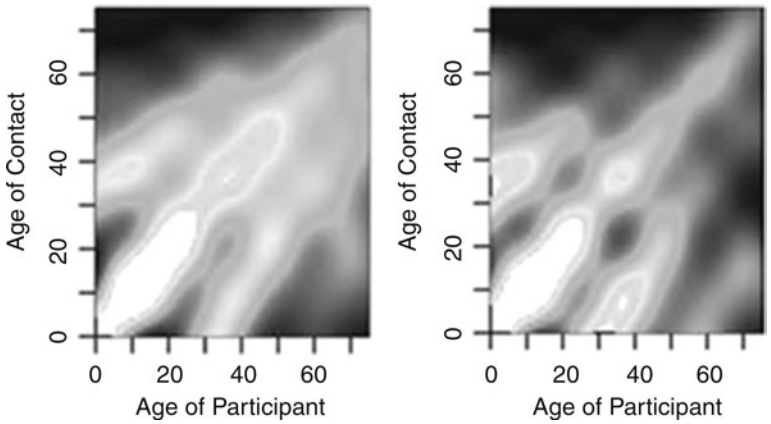

Fig. 5.9 Smoothed contact matrices among age classes in Great Britain, including all contacts (*left*) or physical contacts only (*right*) (Modified from Mossong et al. 2008). White indicates *high* contact rates, while *dark grey* stands for *low* contact rates

structure the contact network among individuals (Keeling and Rohani 2008; Mossong et al. 2008, Fig. 5.9). This structure may have spatial (due to environmental structure), behavioural (*e.g.*, related to sexual behaviour) or social components. In all cases, pathogen spread occurs at two scales: local or within-group transmission is related to direct (between individuals) or indirect (*e.g.* because of a shared environment) contacts, while between-group transmission is possible through long-distance individual movements (migration, visits, etc., Barlow et al. 1998).

It is important to understand how these structures affect pathogen invasion, spread and persistence for helping decision making in public and veterinary health (Cross et al. 2005; Grenfell and Harwood 1997; Hagenaars et al. 2004; Keeling and Rohani 2002; Lloyd and Jansen 2004). The most studied aspect is spatial structure, which has been modelled in a variety of ways, considering either continuous or discrete space, and sometimes including real environmental characteristics. One of the most widely used concepts is the metapopulation (described in the next subsection), where space is divided into discrete patches, each patch representing a potential localisation of a group of hosts.

5.3.4 Disease Spread in Metapopulations

A metapopulation structure corresponds to an inter-patches contact network in which space is either implicit or explicit (Fig. 5.10). Each unit corresponds to either an individual (Rhodes and Anderson 1996), or a local population (Cross et al. 2005; Park et al. 2002). The concept of metapopulation has been largely used in ecology and population genetics to study dynamics of fragmented populations and genes

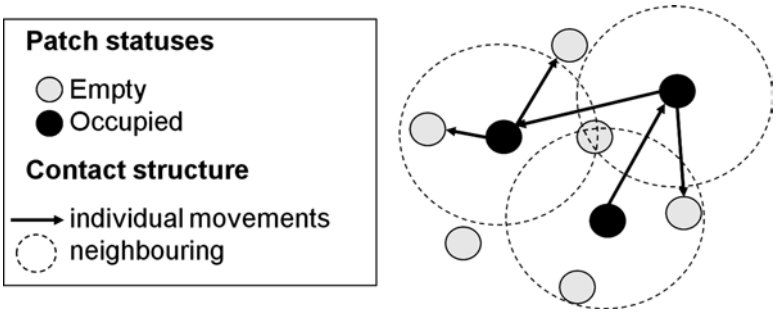

Fig. 5.10 A spatially explicit metapopulation with a contact structure based on individual movements and neighbouring relationships

flow on heterogeneous landscapes (Grenfell and Harwood 1997). Numerous epidemiological models have been developed in a context of metapopulation, accounting for patch infection dynamics in addition to the extinction-colonisation process. Hess (1996) has been a precursor by conceptualising an epidemiological model based on Levin's model: the migration of infectious individuals is a source of infection for susceptible individuals from pathogen-free populations.

Infection spread between patches has been modelled either mechanistically, *i.e.*, explicitly representing the phenomena that are at the origin of disease transmission (Cross et al. 2005; Jesse et al. 2008; Keeling and Rohani 2002), or phenomenologically, by considering that the presence of infection in a patch results in a positive force of infection on patches in contact without explicitly representing the transmission process between patches (Hagenaars et al. 2004; Keeling and Rohani 2002; Park et al. 2002).

From a mechanistic point of view, three types of metapopulations can be defined based on the type of contacts between patches (for a review, see Keeling and Rohani 2008). First, individuals do not encounter explicitly but indirect contacts exist because of the wind, a mobile reservoir, a vector, or through neighbouring contacts between adjacent populations. This mechanism is well adapted to plants which are static, but also to cases when animal diseases are vectored among herds, such as in the case of foot-and-mouth virus aerial transmission among herds, and to neighbouring populations sharing a common environment (water point, feeding area, etc.). Second, individuals may explicitly move between patches with no return to their source patch, as it is the case when individuals disperse, are sold or bought. Third, individuals may move between patches and then return to their source patch (human populations or seasonal migration). In this last case, the duration of the visit influences the number of cases generated by a visitor if infected or the probability for a susceptible visitor of being infected.

The contact pattern may be represented by a contact matrix among patches. For homogeneous networks, all patches are equally connected to each other (Hagenaars et al. 2004; Jesse et al. 2008). For heterogeneous networks, a contact matrix defines which patches are in contact (Cross et al. 2005; Park et al. 2002). In a simplistic

approach, the intensity of contacts may be equivalent for all couples of patches, whereas more refined models consider variable contact rates among patches, using observed or modelled contact networks.

Last, metapopulation models may have various levels of complexity, depending on whether they account for the within-patch infection dynamics. Metapopulations may first have no explicit within-patch dynamics (each patch is considered to have a global infection status). Such an approach has been used to study the persistence of a metapopulation when infected by a pathogen (Gog et al. 2002; Hess 1996; McCallum and Dobson 2002) or pathogen spread and persistence when local infection dynamics rapidly reaches an equilibrium (*e.g.*, avian flu: Le Menach et al. 2006; foot-and-mouth disease: Le Menach et al. 2005). When the within-patch infection dynamics is modelled, the infection status of each individual is considered (Cross et al. 2005; Hess 1996; Jesse et al. 2008; Park et al. 2002) and the status of patches is derived from patch composition. This approach is useful when there is a high variability in the within-patch prevalence among infected patches or for a given patch over time. For example, considering bovine paratuberculosis, infected animals may exit the herd long before being infectious because of a long latency period between infection and shedding (Marcé et al. 2011). These models are more realistic and give a better overview of all possible epidemiological situations, but are also far more complex.

Representing or not the within patch dynamics depends, in addition to the question under study, on the separation between time-scales of processes occurring within and between patches. If local dynamics are fast and global dynamics are slow, it is possible to neglect the first ones under certain stability assumptions and thus to reduce complexity.

5.4 Models for Evaluating Control Strategies of Pathogen Spread in a Population

Providing help guide for decision making about diseases spread prevention and control is one of the major purposes of epidemiological modelling. The objective of such interventions is to prevent emergence and to reduce the incidence in new cases and hence the total epidemic burden. More generally, this aims at optimizing economic animal or human-health outcomes. If we consider the case of animal infectious diseases, their control relies on three principles: increasing resistance in infection of susceptible animals (*e.g.* through vaccination or genetic selection), reducing or preventing shedding of the pathogen by infectious animals (through treatment, test-and-cull strategies), or preventing contacts between susceptible animals and pathogens (through confinement, quarantine, movement restrictions) (Garner et al. 2007). Disease control may also involve indirect measures such as acting on population dynamics (through culling, contraception, modified hunting strategy or renewal strategy in a herd) or acting on the environment (*e.g.* through sanitary fences, Ward et al. 2009). Each strategy is based on a single or a combination of measures, which may be implemented at different levels and scales.

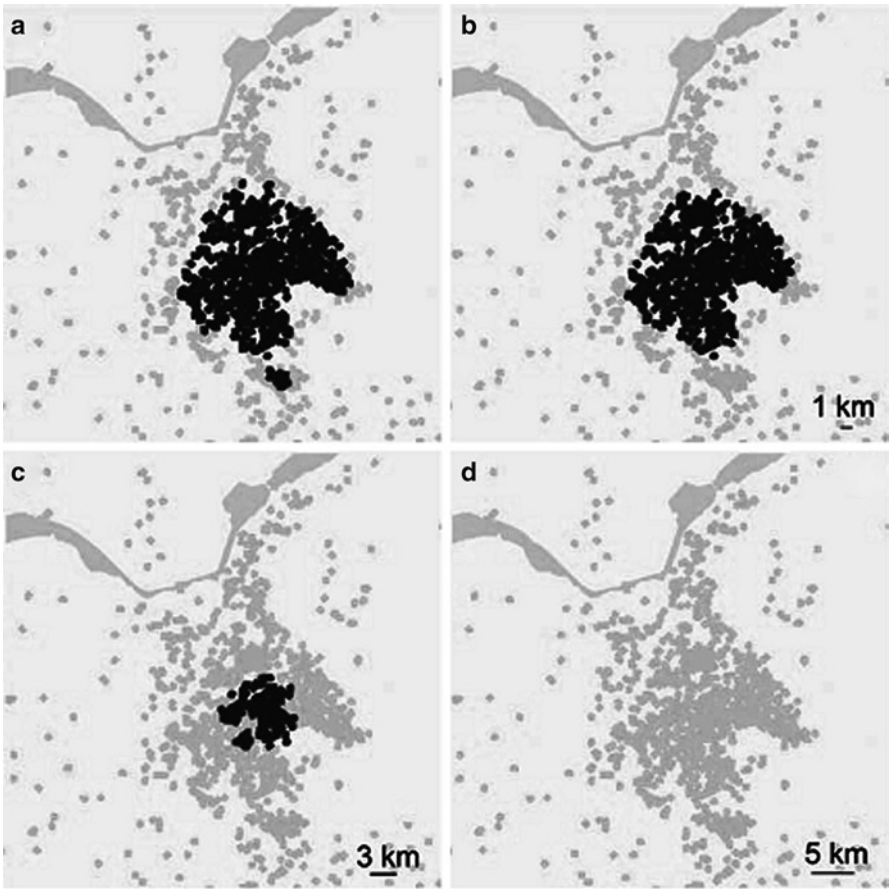

Fig. 5.11 High-risk areas for epidemic spread of the H7N7 avian influenza in poultry for various local culling strategies in the Netherlands (Modified from Boender et al. 2007). (**a**) Results for the default scenario (no culling); (**b**) Results for a scenario with immediate culling of all farms within a range of 1 km around an infected farm; (**c**, **d**) Culling is carried out in a range of 3 and 5 km around infected farms, respectively. Farms in *light grey* pose no risk of epidemic spread for the chosen control strategy, while farms in *dark grey* constitute a risk of epidemic spread even in the presence of interventions

Modelling is an adequate tool for comparing, implementing, evaluating and optimizing control strategies, by allowing to test *ex-ante* a large number of scenarios, resulting from the multiplicity of measures that can be combined and from the interaction between disease spread and population dynamics. Questions to be answered are numerous and various: what size for the zone of preventive vaccination? Which animals should be targeted by tests, vaccination?

Models – for which the consistence between outputs and expectations has been evaluated – can then be used to qualitatively compare control strategies.

For example, Wilkinson et al. (2004) compared various vaccination strategies designed to limit the transmission of bovine tuberculosis in badgers, such as proactive vaccination versus vaccination in reaction to cattle infection, or large-scale versus localized implementation. Models that have been validated on data can be used to provide predictions on the effectiveness of interventions, at least comparatively. Boender et al. (2007, Fig. 5.11) used data from a recent epidemic of avian influenza H7N7 in the Netherlands to fit an explicitly spatial model. Various levels of culling were then tested, to examine the balance between cost and effectiveness of culling.

However, using quantitative predictions requires accurate parameterisation of models and thorough validation of their forecasts using appropriate data sets which are not always recorded. In particular, when a new pathogen emerges, no historical data are available. Moreover, a reference situation in the absence of any intervention or with a perfectly known control strategy is rarely described, especially for endemic diseases. Therefore, the use of modelling approaches to evaluate control strategies should be preferentially considered for qualitative assessment of their impact.

5.5 Conclusion

Modelling is a powerful tool for representing complex systems, testing hypotheses, estimating key parameters from data and predicting the outcome of host-pathogen interactions without or in the presence of interventions. When knowledge and data are available at different scales, models also allow relating fine scales at which mechanisms are known to larger scales at which observations can be made, in order for instance to estimate parameter values (Soubeyrand et al. 2007).

Any model involves a trade-off between simplicity and mathematical tractability on one hand and complexity allowing a closer similarity to the specific problem under study on the other. In this process, taking into account the genetic and molecular variability of both hosts and pathogens is the coming challenge (Anderson 1995; Galvani 2003). Several attempts have been made to integrate the genetic diversity of strains, or represent simple selective processes. For instance, taking into account both epidemiological and molecular relationships between infected premises allowed Cottam et al. (2008) to trace back the spatio-temporal, as well as the evolutionary history of a beginning foot-and-mouth epidemic (see the following chapters). For influenza viruses, several models have been built to analyze the interaction between the pattern of reinfection and the mutation process (Gordo et al. 2009). However, classical tools of epidemiological modelling may not be relevant when genetic variability is involved. The development of specific tools is required to integrate evolutionary ecology and epidemiological patterns.

References

Anderson RM (1995) Evolutionary pressures in the spread and persistence of infectious agents in vertebrate populations. Parasitology 111:S15–S31

Anderson RM, May RM (1991) Infectious diseases of humans: dynamics and control. Oxford University Press, Oxford

Anderson RM, May RM, McLean AR (1988) Possible demographic consequences of AIDS in developing countries. Nature 332:228–234

Andersson H, Britton T (2000) Stochastic epidemic models and their statistical analysis, Springer lecture notes in statistics, vol 151. Springer, New York

Bacaër N, Gernaoui S (2006) The epidemic threshold of vector-borne diseases with seasonality. J Math Biol 53:421–436

Ball F, Lyne OD (2001) Stochastic multi-type SIR epidemics among a population partitioned into households. Adv Appl Prob 33:99–123

Barlow ND, Kean JM, Caldwell NP, Ryan TJ (1998) Modelling the regional dynamics and management of bovine tuberculosis in New Zealand cattle herds. Prevnt Vet Med 36:25–38

Becker N (1979) The uses of epidemic models. Biometrics 35:298–305

Becker NG (1989) Analysis of infectious disease data. Chapman & Hall, London

Bernoulli D (1766) Essai d'une nouvelle analyse de la mortalité cause par la petite vérole, et des avantages de l'inoculation pour la prévenir. Histoire de l'Académie royale des sciences: mémoires de mathématiques et de physiques, pp 1–40

Boender GJ, Hagenaars TJ, Bouma A et al (2007) Risk maps for the spread of highly pathogenic avian influenza in poultry. PLoS Comput Biol 3:e71. doi:10.1371/journal.pcbi.0030071

Cauchemez S, Boëlle PY, Donnelly CA et al (2006) Real-time estimates in early detection of SARS. Emerg Infect Dis 12:110–113

Choisy M, Rohani P (2006) Harvesting can increase severity of wildlife disease epidemics. Proc Roy Soc B 273:2025–2034

Cottam EM, Thébaud G, Wadsworth J et al (2008) Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. Proc Biol Sci 275:887–895

Cross PC, Lloyd-Smith JO, Johnson PLF, Getz WM (2005) Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. Ecol Lett 8:587–595

Daley DJ, Gani J (1999) Epidemic modeling. Cambridge University Press, Cambridge

Diekmann O, Heesterbeek JAP (2000) Mathematical epidemiology of infectious diseases: model building, analysis and interpretation. Wiley, New York

Diekmann O, Heesterbeek JAP, Metz JAJ (1990) On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J Math Biol 28:365–382

Ezanno P, Fourichon C, Viet AC, Seegers H (2007) Sensitivity analysis to identify key-parameters in modelling the spread of bovine viral diarrhoea virus in a dairy herd. Prevent Vet Med 80:49–64

Ezanno P, Fourichon C, Seegers H (2008) Influence of herd structure and type of virus introduction on the spread of bovine viral diarrhoea virus (BVDV) within a dairy herd. Vet Res 39:39

Ferguson NM, Donnelly CA, Anderson RM (2001) The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science 292:1155–1160

Ferguson NM, Cummings DA, Fraser C et al (2006) Strategies for mitigating an influenza pandemic. Nature 442:448–452

Ferrandino FJ (2008) Effect of crop growth and canopy filtration on the dynamics of plant disease epidemics spread by aerially dispersed spores. Phytopathology 98:492–503

Flahault A, Vergu E, Boëlle PY (2009) Potential for a global dynamic of Influenza A (H1N1). BMC Infect Dis 9:129

Galvani AP (2003) Epidemiology meets evolutionary ecology. Trends Ecol Evol 18:132–139

Garner MG, Dubé C, Stevenson MA et al (2007) Evaluating alternative approaches to managing animal disease outbreaks – the role of modelling in policy formulation. Vet Italiana 43:285–298

Gerbier G, Bacro JN, Pouillot R et al (2002) A point pattern model of the spread of foot-and-mouth disease. Prevent Vet Med 56:33–49
Glass K, Grenfell BT (2003) Antibody dynamics in childhood diseases: waning and boosting of immunity and the impact of vaccination. J Theor Biol 221:121–131
Gog J, Woodroffe R, Swinton J (2002) Disease in endangered metapopulations: the importance of alternative hosts. Proc R Soc B 269:671–676
Gordo I, Gomes MGM, Reis DG, Campos PRA (2009) Genetic diversity in the SIR model of pathogen evolution. PLoS One 4:e4876. doi:10.1371/journal.pone.0004876
Greenhalgh D, Diekmann O, De Jong MCM (2000) Subcritical endemic steady states in mathematical models for animals infections with incomplete immunity. Math Biosci 165:1–25
Grenfell B, Harwood J (1997) (Meta) population dynamics of infectious diseases. Trends Ecol Evol 12:395–399
Greenwood M (1931) On the statistical measure of infectiousness. J Hyg (Lond) 31:336–351
Hagenaars TJ, Donnelly CA, Ferguson NM (2004) Spatial heterogeneity and the persistence of infectious diseases. J Theor Biol 229:349–359
Hamer WH (1906) Epidemic diseases in England – the evidence of variability and of persistency of type. Lancet 1:733–739
Heesterbeek JAP (2002) A brief history of R_0 and a recipe for its calculation. Acta Biotheor 50:189–204
Heesterbeek JAP, Roberts MG (2007) The type-reproduction number T in models for infectious disease control. Math Biosci 206:3–10
Heffernan JM, Smith RJ, Wahl LM (2005) Perspective on the basic reproductive number. J R Soc Interface 2:281–293
Hess G (1996) Disease in metapopulation models: implications for conservation. Ecology 77:1617–1632
Hethcote HW (2000) The mathematics of infectious diseases. SIAM review 42:599–653
Jesse M, Ezanno P, Davis S, Heesterbeek JAP (2008) A fully coupled, mechanistic model for infectious disease dynamics in a metapopulation: movements and epidemic duration. J Theor Biol 254:331–338
Keeling MJ, Rohani P (2002) Estimating spatial coupling in epidemiological systems: a mechanistic approach. Ecol Lett 5:20–29
Keeling MJ, Rohani P (2008) Modeling infectious diseases in humans and animals, 1st edn. Princeton University Press, Princeton
Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. Proc R Soc A 115:700–721
Kernéis S, Grais RF, Boëlle PY et al (2008) Does the effectiveness of control measures depend on the influenza pandemic profile? PLoS One 3:e1478
Le Menach A, Legrand J, Grais RF et al (2005) Modeling spatial and temporal transmission of foot-and-mouth disease in France: identification of high-risk areas. Vet Res 36:699–712
Le Menach A, Vergu E, Grais RF et al (2006) Key strategies for reducing spread of avian influenza among commercial poultry holdings: lessons for transmission to humans. Proc Biol Sci 273:2467–2475
Lloyd AL, Jansen VAA (2004) Spatiotemporal dynamics of epidemics: synchrony in metapopulation models. Math Biosci 188:1–16
Lloyd-Smith JO, Cross PC, Briggs CJ et al (2005a) Should we expect population thresholds for wildlife disease? Trends Ecol Evol 20:511–519
Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005b) Superspreading and the effect of individual on variation disease emergence. Nature 438:355–359
Longini IM Jr, Nizam A, Xu S et al (2005) Containing pandemic influenza at the source. Science 309:1083–1087
Marcé C, Ezanno P, Seegers H, Pfeiffer D, Fourichon C (2011) Predicting fadeout versus persistence of paratuberculosis in a dairy cattle herd for management and control purposes: a modelling study. Vet Res 42:36. doi:10.1186/1297-9716-42-36
Matthews L, Low JC, Gally DL et al (2006) Heterogeneous shedding of *Escherichia coli* O157 in cattle and its implications for control. PNAS 103:547–552

McCallum H, Dobson A (2002) Disease, habitat fragmentation and conservation. Proc R Soc B 269:2041–2049
Mossong J, Hens N, Jit M et al (2008) Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med 5(3):e74. doi:10. 1371/journal.pmed.0050074
Murphy BM, Singer BH, Kirschner D (2003) On treatment of tuberculosis in heterogeneous populations. J Theor Biol 223:391–404
Novozhilov AS (2008) On the spread of epidemics in a closed heterogeneous population. Math Biosci 215:177–185
Park AW, Gubbins S, Gilligan CA (2002) Extinction times for closed epidemics: the effects of host spatial structure. Ecol Lett 5:747–755
Real LA, Childs JE (2006) Spatial-temporal dynamics of rabies in ecological communities. In: Collinge SK, Ray C (eds.) Disease ecology – community structure and pathogen dynamics. Oxford University Press, Oxford
Restif O, Grenfell BT (2007) Vaccination and the dynamics of immune evasion. J R Soc Interface 4:143–153
Rhodes CJ, Anderson RM (1996) Persistence and dynamics in lattice models of epidemic spread. J Theor Biol 180:125–133
Riley S, Fraser C, Donnelly CA et al (2003) Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science 300:1961–1966
Roberts MG (2007) The pluses and minuses of R_0. J R Soc Interface 4:949–961
Roberts MG, Heesterbeek JAP (2003) A new method for estimating the effort required to control an infectious disease. Proc R Soc B 270:1359–1364
Roberts MG, Heesterbeek JAP (2007) Model-consistent estimation of the basic reproduction number from the incidence of an emerging infection. J Math Biol 55:803–816
Rosà R, Pugliese A, Villani A, Rizzoli A (2003) Individual-based vs. deterministic models for macroparasites: host cycles and extinction. Theor Popul Biol 63:295–307
Ross R (1908) Report on the prevention of malaria in Mauritius. Waterlow & Sons, London
Rüegg SR, Heinzmann D, Barbour AD, Torgerson PR (2008) Estimating the transmission dynamics of *Theileria equi* and *Babesia caballi* in horses. Parasitology 135:555–565
Saltelli A, Chan K, Scott M (2000) Sensitivity analysis. Wiley series in probability and statistics, Wiley, New York
Soubeyrand S, Thébaud G, Chadoeuf J (2007) Accounting for biological variability and sampling scale: a multi-scale approach to building epidemic models. J R Soc Interface 4:985–997
Svensson A (2007) A note on generation times in epidemic models. Math Biosci 2087:300–311
Thrusfield M (1995) Veterinary epidemiology, 2nd edn. Blackwell Science, London
Valleron AJ (2000) Les rôles de la modélisation en épidémiologie. C R Acad Sci Paris, Sciences de la vie 323:429–433
van den Driessche P, Watmough J (2002) Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math Biosci 180:29–48
Veliov VM (2005) On the effect of population heterogeneity on dynamics of epidemic diseases. J Math Biol 51:123–143
Wallinga J, Lipsitch M (2007) How generation intervals shape the relationship between growth rate and reproductive numbers. Proc R Soc B 274:599–604
Wallinga J, Teunis P (2004) Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. Am J Epidemiol 160:509–516
Ward AI, VerCauteren KC, Walter WD et al (2009) Options for the control of disease 3: targeting the environment. In: Delahay RJ, Smith GC, Hutchings MR (eds.) Management of disease in wild mammals. Blackwell Science, Oxford
Wilkinson D, Smith GC, Delahay RJ, Cheeseman CL (2004) A model of bovine tuberculosis in the badger *Meles meles*: an evaluation of different vaccination strategies. J Appl Ecol 41:492–501
Yan P (2008) Separates roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. J Theor Biol 251:238–252

Chapter 6
Evolution of Virulence: Intuitions and Models

Jean-Baptiste Ferdy and Sylvain Gandon

Abstract In this chapter, we review some of the classic theoretical results on virulence evolution in pathogens. Our goal is to contrast an intuitive vision of this question to results that can be obtained using simple modeling tools. The main conclusion of this chapter could be that, in pathogens as in many other natural systems, natural selection does not maximize the basic reproductive rate R_0. We illustrate this point with several scenarios that take into account several complexities such as superinfection and vertical transmission. We conclude this chapter with some considerations on how the use of these simple models can shed light on some experiments and, in return, how some experiment data is required to improve our theoretical predictions on pathogen evolution.

6.1 Introduction

In this chapter we will review some of the classical results produced by theoreticians on the evolutionary biology of pathogens. The purpose here is not to give a complete account of the mathematical tools that are required for this type of work – although we hope that reading this chapter can constitute a first step in learning those.[1]

[1] The reader who would like to go in more details into the mathematical techniques we use in this chapter should read (Dieckmann et al. 2002; Feng et al. 2006; Day and Gandon 2006, 2007).

J.-B. Ferdy (✉)
Laboratoire Évolution et Diversité Biologique, Université Toulouse 3, CNRS UMR 5174,
118 route de Narbonne, 31062 Toulouse Cedex 9,
e-mail: ferdy@cict.fr

S. Gandon
Centre d'Écologie Fonctionnelle et Évolutive, CNRS UMR 5175,
1919 route de Mende, 34293 Montpellier
e-mail: sylvain.gandon@cefe.cnrs.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_6,

We rather tried to contrast intuitions that most biologists have on how virulence should evolve and what theory really says about it. The main message of this chapter could be summarized by saying that natural selection does not always maximize the basic reproduction ratio, R_0, of the parasite. Of course this is not a recent discovery and this message could apply to many different biological systems, other than pathogens (the R_0 is often called the lifetime reproductive success in classical life-history theory see for example Roff 1992; Stearns 1992). But we think that some of the consequences of this general rule are sufficiently counter-intuitive to deserve a careful inspection in the context of pathogen evolution.

In the following we describe three simple models for which we determine the optimal virulence, that maximizes R_0. We then contrast this analysis to a more rigorous study of the selective pressures that play on pathogens. Finally we conclude by presenting a recent modeling technique that allow to describe transient pathogen evolution in non-equilibrium populations. As a conclusion, we present the case of polymorphism in pathogen population where we think that a good understanding of the theory of pathogen evolution is required for a correct interpretation of experimental results.

6.2 Modeling Epidemics

Before we can study how parasites and pathogens evolve, we first need to build some mathematical description of how they propagate in a population of hosts. Let us define s as the density of susceptible hosts in a population. Let i be the density of infected individuals in the same population. Classically, mathematical models will relate variation in those densities (ds or di) over a short period of time (dt) to fundamental biological processes such as reproduction, death and infection. The simple model we want to define here is represented by Fig. 6.1.

In this model, susceptible hosts reproduce at a rate $f(n)$ that decreases with the total number of individuals ($n = s + i$) in the population. This decrease corresponds to the fact that hosts depend on a limited resource for their reproduction. We do not need here to specify in more details how the function f describes the effects of competition on reproduction. Infected individuals also reproduce in this model, but their fecundity, relative to that of susceptible hosts, is φ. As for reproduction, of death occurs for all hosts. But infected individuals die at a rate $\mu + \alpha$ while susceptible hosts die at a rate μ. The additional mortality α that is caused by infection is therefore a way to measure the virulence of pathogens. Finally, infected individuals contaminate susceptible ones by contact. We suppose here that the rate of contact between susceptible and infected individuals is proportional to the product of their density. The parameter β is thus a measure of the ability of the pathogen to contaminate a susceptible individual when it has encountered an infected one.

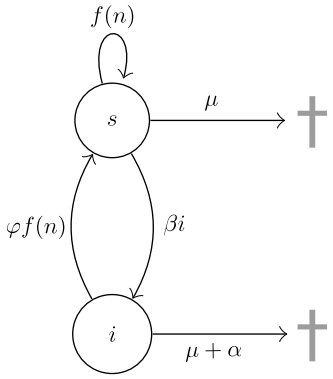

Fig. 6.1 An epidemiological model where pathogens are transmitted horizontally, by contact between susceptible and infected hosts. The variable *n* is the total number of hosts in the population: $n = s + i$. The function *f*(*n*) describes the number of offspring produced by a susceptible host per unit of time. This function decreases with *n*, because hosts depends on a limited resource for their reproduction. The parameter φ is the fecundity of infected individuals relative to that of susceptible ones. The parameter μ is the death rate of susceptible hosts, and α is the additional mortality that infected individuals experience because of the pathogen virulence. Finally, susceptible host are contaminated at a rate βi, the number of new infections produced per unit time being thus βsi

Putting all these terms together, we obtain the following set of differential equations:

$$\begin{cases} \dfrac{ds}{dt} = & (s + \varphi i) f(n) - \beta si - \mu s \\ \dfrac{di}{dt} = & i\big(\beta s - (\mu + \alpha)\big) \end{cases} \tag{6.1}$$

These equations describe the variation of *s* and *i* over time. They can help understand the fundamental characteristics that the pathogen must possess to be able to develop an epidemic. More precisely, *i* will increase over time if *di/dt* is positive. The sign of this growth rate depends on the rate horizontal transmission, βs, which increases with the number of hosts that can potentially be infected, and on the death rate $\mu+\alpha$.

Let us imagine an initial situation where the pathogen is absent from the host population. In this circumstances, we have $i = 0$ and $s = \hat{s}_{\varnothing}$ such that $f(\hat{s}_{\varnothing}) = \mu$. Let us imagine now that a single infected individual is introduced in this population. The number of infected individuals will increase from this unique initial infection if

$$\beta \hat{s}_{\varnothing} - (\mu + \alpha) > 0 \tag{6.2}$$

Let us now consider the same situation and use, as most often done, the R_0 parameter to predict whether the pathogen initiates an epidemics. This parameter R_0 is defined as the number of secondary infections that are formed from a unique initial

infection. To compute this quantity, we thus have to multiply the number of infections that are produced by an infected individual per unit of time, which is $\beta \hat{s}_{\varnothing}$ as we saw before, by the life span of this infected individual. The death rate of infected individuals is $\mu+\alpha$. Their life span can thus be estimated as $1/(\mu+\alpha)$. Finally, R_0 can therefore be computed as

$$R_0 = \frac{\beta \hat{s}_{\varnothing}}{\mu+\alpha} \tag{6.3}$$

and the epidemic will develop if R_0 is greater than one, which is strictly equivalent to condition (6.2).

6.3 A First Approach to Pathogen Evolution: The Maximization of R_0

We saw in the previous section that pathogens can persist in a population of host if $R_0 > 1$. The quantity R_0 can thus be viewed as a measure of epidemic potential of a pathogen in a fully susceptible population. This quantity is often used as a measure of pathogen's fitness. Under this hypothesis natural selection should favor pathogen strains with high transmission rates and low virulence. Indeed R_0 increases with the transmission rate and the expected duration of the infection. Yet, maximizing transmission and minimizing virulence simultaneously is likely to be very difficult for pathogens. Transmission often relies on host exploitation and thus any increase in the capacity to contaminate susceptible hosts would come at the expense of an increased virulence (see Alizon et al. 2008, for a review). The left panel in Fig. 6.2 illustrates a possible relationship between transmissibility (β) and virulence (α). The right panel of the same figure presents R_0, the measure of transmission presented in equation (6.3), when virulence varies. When virulence is low, R_0 is low because pathogens are not contagious enough. When virulence is high, R_0 is also low, but this is this time because infected hosts die rapidly after infection, which reduces the opportunities of contamination. Because the transmissibility of pathogens and the lifespan of the hosts it infects are both connected to virulence, transmission is maximized for an intermediate value of virulence (α^{opt} in Fig. 6.2).

In the previous paragraph, we supposed that natural selection should bring pathogens toward α^{opt}. This prediction relies on the assumption that natural selection maximizes R_0. A possible flaw in this hypothesis is that R_0 is measured in a context where only susceptible hosts are present. But in nature pathogens evolve by a combination of mutation, which creates strains with new virulence and transmission capacities, and competition between pathogenic strains. The population of pathogens will evolve if a mutant strain, that is initially rare, is capable to increase in frequency. What matters for the mutant strain, therefore, is to be better transmitted than other strains that are present in the population of hosts. The parameter R_0 has not been built to measure this capacity.

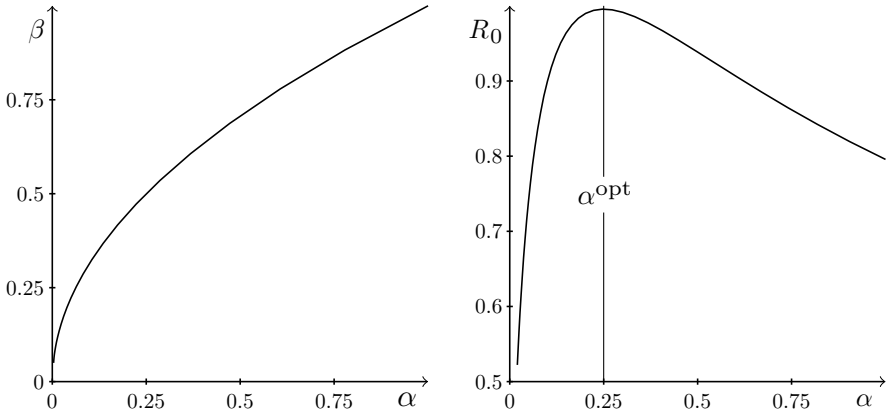

Fig. 6.2 *Left panel*: the curve presents an hypothetical relationship between virulence (α) and transmissibility (β) in pathogens. This type of relationship assumes that β increases with α but at a diminishing rate (i.e. slower than linearly: compare curve to straight line). This type of relationship is often assumed because it corresponds to the idea that it is more difficult, and thus costly, to increase transmissibility for pathogens that are already very contagious than for those that are not. *Right panel*: the induced relationship between virulence (α) and transmission (R_0). Because β increases slower than α, R_0 is maximized for an intermediate value of virulence (see equation (5.3)

The situation we described above can be modeled as a case of competition between a rare mutant strain (with virulence α_m and transmissibility β_m) and a resident strain (with virulence α_r and transmissibility β_r). This is done by adding a third differential equation to those of Eq. (6.1), which describes the dynamics of the population of hosts infested by the mutant strain:

$$\frac{di_m}{dt} = i_m\left(\beta_m s - (\mu + \alpha_m)\right) \tag{6.4}$$

From there, the condition for mutant invasion can be derived as

$$w(\alpha_m, \alpha_r) = \frac{\beta_m \hat{s}_r}{\mu + \alpha_m} > 1 \tag{6.5}$$

where $w(\alpha_m, \alpha_r)$ can be understood as a measure of fitness of the mutant strain in the context of a population that harbors only the resident pathogenic strain, and where $\hat{s}_r$ is the equilibrium density of susceptible hosts in this population. This condition for invasion is classically represented, as in Fig. 6.3, by a pairwise invisibility plot. This graph represents the set of mutants that can invade in a given resident population. It also allows to predict how virulence should respond to natural selection. Indeed, in the left part of Fig. 6.3, the only mutants that can invade in the resident population are above the diagonal. They are thus more virulent than the resident strain and selection should therefore induce an increase in virulence. Conversely, in the right part of the same graph, mutants that can invade are all below the diagonal. Selection should then induce a decrease in virulence. As a result, selection should

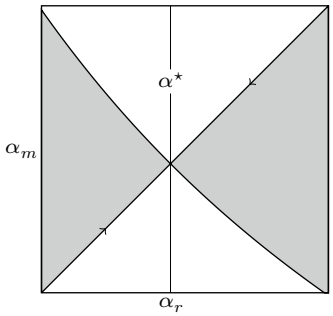

Fig. 6.3 A pairwise invisibility plot (PIP). For each resident pathogen with virulence α_r in its equilibrium population, the fate of a mutant with virulence α_m is indicated by the color. Mutants in the *gray area* can invade in the resident population. Conversely, mutants in the *white area* are counter selected

make α evolve toward the value α^* which lies at the center of the graph. Once this point is reached, it is clear from Fig. 6.3 that no mutant is able to invade the resident population. The point α^* is therefore an Evolutionary Stable Strategy (ESS).

We shall now compare the value α^*, that we identified in Fig 6.3, to the value α^{opt} that maximizes R_0. From Eq. (6.1), it can easily be shown that $w(\alpha_r, \alpha_r) = (\beta_r \hat{s}_r)/(\mu + \alpha_r) = 1$ so that Eq. (6.5) can be rewritten as

$$\frac{\beta_m}{\mu + \alpha_m} > \frac{\beta_r}{\mu + \alpha_r} \tag{6.6}$$

Multiplying both sides of this equation by $\hat{s}_\varnothing$, we obtain two terms that are identical to Eq. (6.3), i.e., to R_0. Eq. 6.6 therefore confirms the fact that the unbeatable strategy α^* is the one that maximizes R_0, as we first hypothesized. Simulations where pathogen variants are introduced by mutation and compete to infect new hosts also confirm this prediction: selection does indeed drive the population of pathogens toward a evolutionary stable (ES) virulence α^* which is identical to α^{opt} and does therefore maximize R_0 (see Fig 6.4).

One might now ask how selection on virulence is affected by environmental changes. For example, what would be the consequences of an increase mortality in hosts? A widespread argument is used to answer this question: pathogens should transmit rapidly to avoid the death of their hosts. Natural selection should thus favor higher virulence in low quality environment than in high quality ones. In the framework we developed above, this prediction can be rephrased as follows: any increase in μ should induce an increase in α^*. This prediction seems to be confirmed by the results presented in Fig 6.5.

We shall now inspect more carefully this result. From Eq. (6.5) and with the assumption that mutations have small effects, we can determine whether mutants that can invade the resident population are more virulent than the resident strain (i.e., above the diagonal in Fig 6.3) or less virulent that the resident strain (i.e., below

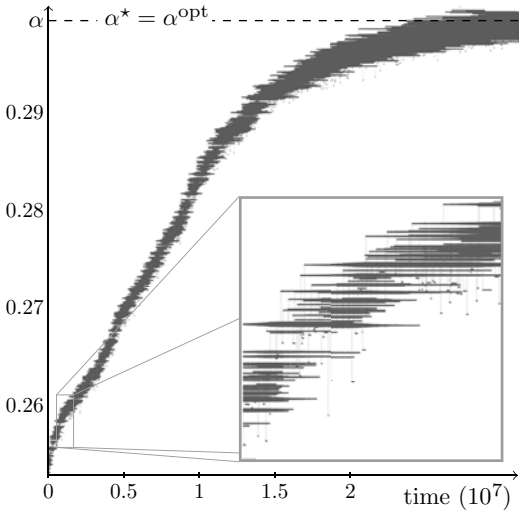

Fig. 6.4 The evolution of α in a model with horizontal transmission only. *Horizontal lines* represent strains of pathogens, the frequency of each strain being represented by the thickness of the line. Thin *vertical lines* trace the origin of each strain by joining the virulence of each strain to that of the strain it originated from. Over time, the population of pathogens evolves toward a value of virulence α^*. It can be shown that this value also maximizes transmission ($\alpha^* = \alpha^{opt}$, see text and Fig. 5.2). The *frame* presents details of this evolution. It allows, in particular, to see that early in the simulation all mutant strains that are less virulent than average rapidly disappear, because they do not transmit as well as other strains of the population. This counter selection of the least transmitted strains yield an increase of average virulence over time. Later in the simulation, once the population of pathogens has reached α^*, selection stabilizes virulence by eliminating strains that are both more and less virulent than average. All these aspects of the evolutionary dynamics can be predicted from the PIP presented in Fig. 5.3.

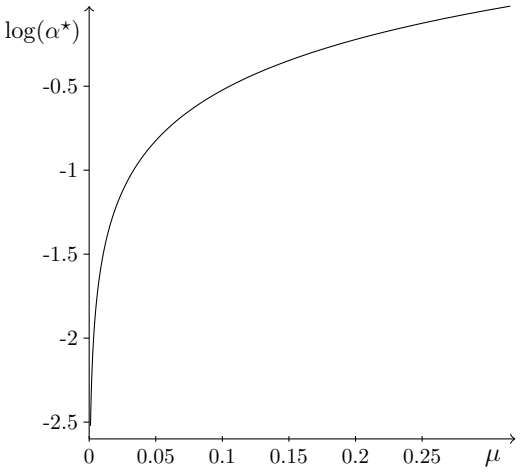

Fig. 6.5 The log-transformed ES virulence α^* as a function of natural host death rate μ

the diagonal in Fig 6.3). In brief, the gradient of selection is obtained from the derivative of the fitness defined in Eq. (6.5) with respect to α_m. It thus relates the effect of a small mutation on the ability of a rare mutant to invade. Here we obtain:

$$\psi(\alpha) = \hat{s}\frac{d\beta}{d\alpha} - 1 \tag{6.7}$$

The first term in $\Psi(\alpha)$ corresponds to the benefit associated to an increase in virulence, in terms of increased contamination opportunities. The second terms quantifies the cost associated with increased virulence, in terms of reduced host life-span. The interpretation of this gradient is thus rather simple: if $\Psi(\alpha) > 0$, selection favors an increase in virulence; if $\Psi(\alpha) < 0$ selection should conversely induce a reduction in virulence. Interestingly, the parameter μ does not appear in the gradient of selection. How should we understand the result presented in Fig. 6.5, then? In fact any increase in μ can easily be shown to provoke an increase in the number of susceptible hosts present in the resident population $(\hat{s})$. An increase in μ therefore opens up new opportunities for transmission, which makes any increase in virulence more beneficial. The verbal prediction we made above was therefore essentially wrong, even though we predicted correctly that α^* should increase with μ.

6.4 Natural Selection Does Not Necessarily Maximize R_0

6.4.1 The Case of Superinfection

In the previous sections we assumed that an infected host cannot be infected by more than a single parasite strain. Taking into account the possibility of multiple infections, however, can dramatically affect the evolution of the parasite. Indeed, allowing for multiple infections adds another level of selection (i.e., within-host selection). In particular, if virulence is associated with some increased competitive ability within the host (de Roode et al. 2004), multiple infections can select for higher levels of virulence (Frank 1996; van Baalen and Sabelis 1995).

The analysis of the effects of multiple infections raises some theoretical challenges as the coexistence between singly and multiply infected hosts generates some host heterogeneity. Although it is possible to take into account this heterogeneity (Gandon 2004), a much simpler alternative relies on the additional assumption that the within-host dynamics is very fast. In this situation, within-host competition leads instantaneously to the success of the resident or the invading strain. In this superinfection model it is no longer necessary to keep track of the dynamics of host infected by more than one strain (Day and Proulx 2004; Gandon et al. 2001; Levin and Pimentel 1981; Nowak and May 1994).

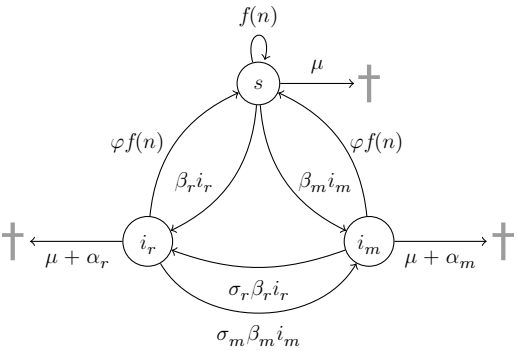

Fig. 6.6 An epidemiological model with superinfection. This model is identical to that described by Fig. 5.1, except that, when two strains of pathogens are present in the population, hosts that are infected by one strain can be superinfected by the other. Here the two strains are denoted *r* (*for resident*) and *m* (*for mutant*). They each have a specific transmissibility (β_r and β_m respectively) and a specific virulence (α_r and α_m respectively). Hosts that are infected by the resident strain (i_r) are super-infected by the mutant strain at a rate $\sigma_m \beta_m i_m$. Symmetrically, hosts that are infected by the mutant strain (i_m) are super-infected by the resident strain at a rate $\sigma_r \beta_r i_r$

Formally, this leads to the following system

$$\begin{cases} \dfrac{ds}{dt} &= \big(s+\varphi(i_r+i_m)\big)f(n)-(\mu+\beta_r i_r+\beta_m i_m)s \\ \dfrac{di_r}{dt} &= \beta_r i_r(s+\sigma_r i_m)-(\mu+\alpha_r+\sigma_m\beta_m i_m)i_r \\ \dfrac{di_m}{dt} &= \beta_m i_m(s+\sigma_m i_r)-(\mu+\alpha_m+\sigma_r\beta_r i_r)i_m \end{cases} \tag{6.8}$$

where σ_r (σ_m) is the probability of superinfection: the probability that a mutant (resident) strain is replaced by a resident (mutant) strain. In the case where i_m is set to zero, superinfection does not occur and again Eq. (6.8) is identical to Eq. (6.1). Similarly, if $\sigma_r = \sigma_m = 0$ the dynamics of the mutant strain is described by Eq. (6.4). This model is represented in Fig. 6.6.

In this model a measure of the fitness of the mutant strain becomes

$$w(\alpha_m,\alpha_r)=\frac{\beta_m(\hat{s}_r+\sigma_m\hat{i}_r)}{\mu+\alpha_m+\sigma_r\beta_r\hat{i}_r} \tag{6.9}$$

Superinfection affects this fitness in two different ways. First, it allows to get access to infected hosts (this is the $\sigma_m \hat{i}_r$ term in the numerator of Eq. (6.9). Second, it reduces the expected duration of an infection by a given strain as, on top of the risk of

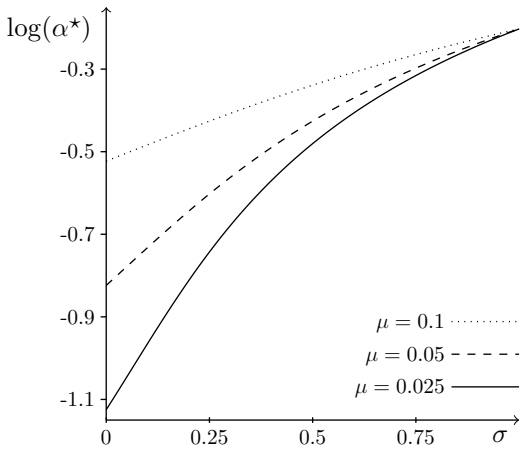

Fig. 6.7 The log-transformed ES virulence of pathogens as a function of σ, the probability of superinfection, and for several values of μ. In these simulations, this probability has been assumed to be the same for the resident and the mutant strain ($\sigma_r = \sigma_m = \sigma$). In all cases, the virulence ES increases with σ. The virulence that maximize R_0 corresponds to the ES virulence only when σ=0

host death, there is also the risk of being out-competed by another strain (this is the $\sigma_r \beta_r \hat{i}_r$ term in the denominator of Eq. (6.9). To better understand the evolutionary consequences of superinfection, we shall now write the gradient of selection, that we introduced in the previous section. We obtain

$$\psi(\alpha) = \left(\hat{s} + \sigma \hat{i}\right) \frac{d\beta}{d\alpha} - 1 \tag{6.10}$$

from where it is clear that the second effect of superinfection that we mentioned above, which should decrease the benefits associated to any gain in virulence, has in fact no impact on the selective process. It is also clear from Eq. (6.10) that superinfection is incorporated in the gradient under the form of an additional benefit in transmission given to any increase in virulence. Superinfection will therefore increase the ES virulence, as illustrated in Fig. 6.7.

As before, we shall now contrast the virulence that would maximize R_0 to the predicted outcome of natural selection. R_0 quantifies the transmission of a single strain of pathogen in the context of a population of fully susceptible hosts. In that context, of course, superinfection cannot occur. Contrary to the fitness of the mutant described by Eq. (6.9), R_0 is therefore not affected by superinfection. It is therefore very clear that the evolutionary consequences of superinfection cannot be predicted from R_0 and consequently that evolution will not maximize R_0 (see Fig. 6.7).

6.4.2 The Case of Vertical Transmission

Let us consider now a more complex case where pathogens are transmitted by a combination of horizontal transmission, as in the previous model, and vertical transmission (see Fig. 6.8). We assume, in this model, that when infected individuals

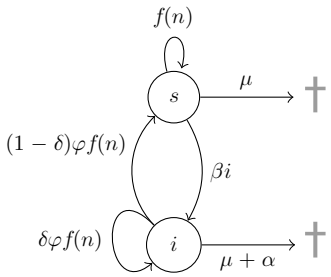

Fig. 6.8 An epidemiological model where pathogens are transmitted both horizontally, by contact between susceptible and infected hosts, and vertically, when infected individuals reproduce. This model is identical to that described in Fig. 6.1, except that a fraction δ of the offspring produced by infected individuals is infected

reproduce, a fraction δ of their offspring is infected while the pathogen is not transmitted to the remaining fraction $1-\delta$. Vertical transmission is therefore imperfect in this model, unless $\delta=1$.

The equations that describe the dynamics of this system can be written as follows:

$$\begin{cases} \dfrac{ds}{dt} = (s+(1-\delta)\varphi i)f(n)-\beta si-\mu s \\ \dfrac{di}{dt} = \delta\varphi if(n)+\beta si-(\mu+\alpha)i \end{cases} \tag{6.11}$$

As above, we can express the condition that a pathogen must satisfy to survive in the population of host as $R_0>1$, with this time

$$R_0 = \frac{\beta \hat{s}_\varnothing}{\mu+\alpha}+\frac{\delta\varphi f(\hat{s}_\varnothing)}{\mu+\alpha} \tag{6.12}$$

This equation is more complex than the previous expression of R_0, but it can conveniently be understood as the sum of two transmission terms. The first term in Eq. (6.12) is indeed identical to Eq. (6.3) and thus quantifies horizontal transmission. The second term is the product of the number of infected offspring produced per unit of time and of the lifespan of an infected individual. It therefore quantifies vertical transmission.

We might now ask, as in the previous section, whether natural selection does maximize R_0. As before, this requires that we consider the fate of a rare mutant pathogenic strain that just appeared in a population where hosts are infected by a resident strain. As before we consider that mutant pathogens have a virulence α_m and a transmissibility β_m, while resident pathogens are characterized by α_r and β_r.

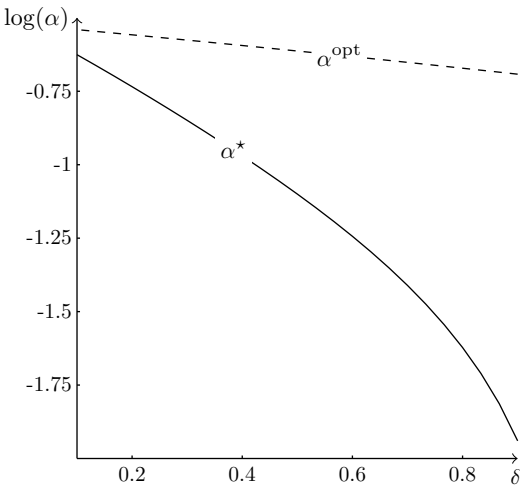

Fig. 6.9 The log transformed optimal (α^{opt}) and evolutionary stable (α^*) virulence in a model with vertical transmission. The virulence α^{opt} maximizes R_0, as defined in Eq. (6.12). The virulence α is the ES virulence, that is the virulence of a resident pathogen that cannot be out-competed by any mutant strain (see Eq. (6.13 and hereafter). The parameter δ controls the efficiency of vertical transmission: infected parents always transmit their pathogen to their offspring when $\delta=1$ and never do it when $\delta=0$. As δ increases both the optimal and the ES virulence decrease, while the discrepancy between these two quantities increases. Natural selection maximizes transmission only in the absence of vertical transmission (i.e., $\alpha^{opt}=\alpha^*$ when $\delta=0$)

The condition that mutants must satisfy to invade in the resident population can be written as

$$w(\alpha_m,\alpha_r)=\frac{\beta_m \hat{s}_r}{\mu+\alpha_m}+\frac{\delta\varphi f(\hat{s}_r+\hat{i}_r)}{\mu+\alpha_m} \tag{6.13}$$

where $\hat{s}_r$ and $\hat{i}_r$ are the equilibrium densities of susceptible and infected individuals in the resident population. As before, it is easily shown that $w(\alpha_r,\alpha_r)=1$ so that the condition for invasion can be written as $w(\alpha_m,\alpha_r)>w(\alpha_r,\alpha_r)$.

The fitness as defined in Eq. (6.13) resembles R_0 as expressed in Eq. (6.12). But contrary to what we saw in Sect. 6.4, $w(\alpha_m,\alpha_r)$ and $w(\alpha_r,\alpha_r)$ cannot both be written as $R_0\hat{s}_r/\hat{s}_\varnothing$. As a result, selection will therefore not maximize R_0, and in fact it is even unclear that some quantity will be maximized over the course of evolution. Still, from Eqs. 6.12 and 6.13, we can numerically compute both the optimal virulence, that maximizes R_0, and the virulence toward which natural selection should bring the population of pathogens. Fig. 6.9 presents a comparison of this optimal virulence (α^{opt}) and of the evolutionary stable (ES) virulence (α^*) that selection favors. From these two calculations, it is quite clear that in this model natural selection does not maximize transmission. In fact the ES virulence is much lower than the optimal virulence, and the difference increases with the efficiency of vertical transmission.

The estimation of transmission that R_0 provides is therefore appropriate to describe the epidemics but not to predict the evolution of the pathogen.

Another surprising outcome of the analysis of this model comes from the calculation of the gradient of selection. From Eq. (6.13) it can be shown that

$$\psi(\alpha) = \hat{s}\frac{d\beta}{d\alpha} - 1 \tag{6.14}$$

which is strictly identical to the gradient of selection we derived for the simple model where pathogens are horizontally transmitted (see Eq. (6.7). Still, Fig. 6.9 clearly shows that the ES virulence sharply decreases when the efficiency of vertical transmission increases. In fact the influence of vertical transmission on selection is, in this model, indirect. It comes through the negative relationship between vertical transmission and the equilibrium density of susceptible individuals: when δ increases, pathogens are transmitted more efficiently and the number of susceptible individuals decreases. As a result the opportunities of horizontal transmission are reduced and selection favors lower values of virulence. This mechanism is similar to the effect of increased host death rate that we described in Sect. 6.3.

6.5 Modeling Transient Evolution

In all the models discussed in Sects. 6.3 and 6.4, the fate of a mutant is considered in a population of resident pathogens that has reached its demographic equilibrium. We shall now discuss a situation where we look at the transient evolution of a parasite population. In contrast with the above sections we will not assume the population to be at an endemic equilibrium but rather to harbor a diversity of strains which are all competing for transmission. To do so we can use the classical formalism of population genetics which has recently been adapted to model the dynamics of the frequency of different pathogen strains, in conjunction with the dynamics of host population size (Day and Gandon 2006; Day and Proulx 2004). The approach begins by first extending the model to allow for k pathogen strains. One then changes variables to model the frequencies of the strains, along with the total number of susceptible and infected individuals. Defining i_T as the total number of infected individuals, and $q_j = i_j / i_T$ as the frequency of strain i, we obtain

$$\begin{cases} \dfrac{ds}{dt} &= (s+\varphi i_T) f(s+i_T) - \overline{\beta}\, s i_T - \mu s \\ \dfrac{di_T}{dt} &= \overline{\beta}\, s i_T - (\mu + \overline{\alpha}) i_T \\ \dfrac{dq_j}{dt} &= q_j (r_j - \overline{r}) \end{cases} \tag{6.15}$$

with

$$r_j = \beta_j s - (\mu + \alpha_j) \tag{6.16}$$

where r_j is a measure of the fitness of strain j, and the overbars denote an expectation over the distribution of pathogen strains. We have assumed here no mutation, no superinfection and no vertical transmission (but see Day and Gandon 2006; Day and Proulx 2004; Day et al. 2007).

In Eq. (6.15), the third differential equation describes the evolutionary dynamics while the two others describe the epidemiological dynamics. At this stage it is possible and often useful to derive a form of Price's equation from Eq. (6.15) that tracks the mean value of any character of interest (Price 1970). For example, if we are interested in the mean level of virulence and transmission, we can differentiate $\bar{\alpha}$ and $\bar{\beta}$ with respect to time, using Eq. (6.15) (Day and Gandon 2006):

$$\begin{pmatrix} \frac{d\bar{\alpha}}{dt} \\ \frac{d\bar{\beta}}{dt} \end{pmatrix} = \mathbf{G} \cdot \begin{pmatrix} -1 \\ s \end{pmatrix} \tag{6.17}$$

where **G** is the genetic (co)variance matrix, i.e., a matrix that comprises both the genetic variances present in the population of pathogens for α and β and the covariance between these traits. This covariance can be understood as a statistical description of the relationship between α and β, replacing the function represented in Fig. 6.2. The vector $(-1, s)$ is the selection gradient and shall thus be compared to Eq. (6.7). Eq. 6.17 is analogous to quantitative genetics models (Day and Proulx 2004; Lande 1976; Lande and Arnold 1983), but it does not rely on Gaussian distributions of strain phenotypes, nor on an assumption of small variance. The product of **G** with the selection gradient in Eq. (6.17) is an exact description of the effect of natural selection on the average level of virulence and transmission. Natural selection favors reduced virulence with a strength of -1. On the other hand, natural selection favors an increased transmission rate with strength proportional to the density of susceptible hosts, s. Thus, for example, direct selection always drives virulence downward at a strength of -1, mediated by the genetic variance in virulence. At the same time, indirect selection pulls virulence upward with a strength proportional to the density of susceptible hosts, mediated by the genetic covariance between transmission and virulence (see Fig. 6.10). An analogous interpretation can be given to the evolutionary dynamics of transmission. In the short term, the **G** matrix and the selection gradient can be assumed constant, and Eq. (6.17) used to predict the direction and the speed of evolution. For longer periods of time, the epidemiological dynamics given in Eq. (6.15) allow to track changes in the selection gradient. In the longer term, changes in the G matrix are more difficult to track since they depend on the selection gradient, the mutation rate, and on the effect of mutations. Additional assumptions regarding the distribution of strain frequencies can be used to derive dynamical equations for these variance components (Day and Proulx 2004).

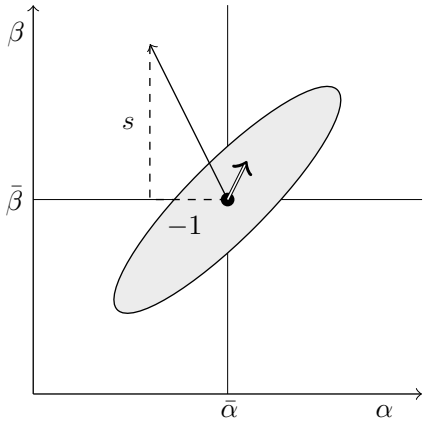

Fig. 6.10 Describing the relation between β and α by a **G** matrix. The *gray ellipse* describes the joint distribution of α and β in a population of pathogen. In this figure, the orientation of the ellipse indicates that any mutation that increase β should also, on average, increase α. The ellipse is therefore a representation of the constraints that couple α and β, as was the function represented in Fig. 5.2. In mathematical terms, the shape of the ellipse can be described by a matrix with the variance in each trait on the diagonal and the covariance between the traits elsewhere. Selection favors both a reduction in α and an increase in β as indicated by the plain arrow. The evolution of the pathogen, indicated here by the double arrow, is determined by the combination of this selective pressure and of the constraints described by the G matrix

6.6 Conclusion

6.6.1 The Different Reasons Why Evolution Cannot Be Predicted from R_0

The quantity R_0 is a convenient measure of the epidemic potential of pathogens. It does describe adequately the initial spread of a pathogen in an otherwise uninfected population of hosts. It predicts the conditions under which a pathogen can form an epidemic. Still, we saw in this chapter that in many situations the R_0 should not be used to predict how pathogens evolve.

The reason why the evolutionary consequences of superinfection cannot be predicted from the R_0 is quite simple to understand: superinfection can occur only when several pathogenic strains co-occur in the same population of hosts. R_0 is estimated in a context where a single strain is present. Quite logically, it does not incorporate the parameters that control the rate of superinfection and, thus, cannot be used to predict anything about how superinfection impact pathogens' evolution.

In the case of vertical transmission, the parameters that describe the efficiency of vertical transmission are all present in the computation of R_0. Still, R_0 measures transmission in a context where most hosts are uninfected. Conversely, the fate of a mutant pathogen strain depends on its capacity to transmit in a context where some

hosts are infected by another pathogenic strains. For this reason, R_0 fails to predict the reduction of ES virulence induced by vertical transmission.

Many other situations of this type could be listed here in which, again, pathogens' transmission as measured by R_0 cannot be taken as a fitness estimate. In fact the first model we present in this chapter should really be seen as an exception. As illustrated above, any deviation from the hypotheses of this model usually makes R_0 useless to predict pathogens' evolution.

6.6.2 *When Natural Selection Maintains Polymorphism in Pathogens Populations*

We just saw that natural selection does not necessarily maximize R_0. In all the cases we detailed in Sect. 6.4, still, natural selection brings pathogens to an intermediate level of virulence that can be thought to maximize some sort of fitness, even though that fitness cannot be estimated directly from the R_0.

In other situations, though, there is no single quantity that natural selection would maximize. This happens, in particular, when several pathogenic strains can coexist in the same population of hosts. In the model we studied in Sect. 6.4.2, for example, polymorphism can arise for certain types of relationship between α and β (Ferdy and Godelle 2005; Lipsitch et al. 1995, 1996). For example a S-shaped relationship, such as that predicted by Alizon and van Baalen (2008), is sufficient to allow the coexistence between vertically and horizontally transmitted pathogens. This illustrates the fact that details on the trade-off functions which are assumed in most models do matter (see Kamo et al. 2007, for another illustration of this particular point). Most past theoretical work has focused on the evolutionary trajectory of pathogens with arbitrary chosen relationships between pathogens life-history traits. Recent work has tried to predict those relationships from basic processes, such as within host pathogen dynamics (Alizon and van Baalen 2005, 2008; Andre et al. 2003; Ganusov and Antia 2003). Yet experimental support for these relationships is still lacking in many host-parasite systems (Alizon et al. 2008).

Assuming that pathogens can be vertically transmitted and that the relationship between α and β is S-shaped, it can be shown that pathogens with different virulence can coexist in the same population of hosts. Fig. 6.11 represents a distribution of α obtained from a simulation of this situation. In these simulations two groups of pathogens coexist. Group a is composed of very virulent and contagious pathogens while group b is composed of pathogens of low virulence which transmission is mostly vertical. The distribution of virulence illustrated in Fig. 6.11 is not stable. It will change over evolutionary time. But the key point here is that in this situation the most frequent pathogens will always be the least virulent.

Imagine now that a biologist samples such a population of pathogens. Most probably he/she would miss the small minority of very virulent pathogens and find only the most frequent, avirulent, strains. As indicated in Fig. 6.11, these strains have a R_0 which is below one. Our hypothetical biologist would then have great difficulties understanding how such strains can survive in nature…In order to interpret

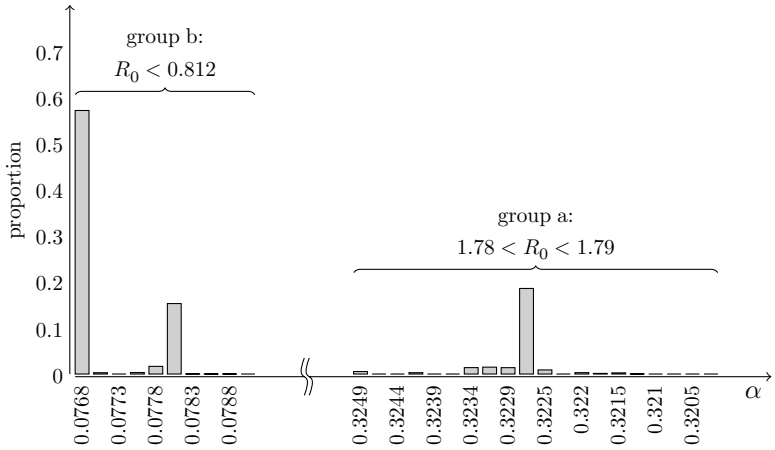

Fig. 6.11 The distribution of α in a polymorphic population of pathogens. The population of pathogens is then composed of two groups which, because of the S-shaped relationship between β and α, coexist in the population of hosts. One group is composed of very contagious and virulent strains (*group a*) while the other is almost incapable to infest susceptible hosts and is transmitted mostly from parent to offspring (*group b*). This second lineage is the most abundant. Surprisingly its R_0 is below one. This means that if those pathogens where introduced alone in a population of susceptible hosts they would not form a viable population. Their vertical transmission would in fact be, in this situation, impeded by the competition for resource with susceptible hosts. When pathogens of the lineage a are present, they infect most susceptible hosts which reduces competition. Vertical transmission is then efficient enough to ensure the survival of pathogens of the lineage b

correctly these findings, it is necessary to recognize that R_0 quantifies invasion potential in a very specific situation that often does not correspond to the natural condition in which these pathogens live and evolve.

This sort of situation could in fact be relatively common in nature, as the coexistence between several pathogenic strains of the same species has been shown to be stable in several other scenarios. For example, polymorphism can be protected in the case of superinfection, if the most virulent pathogens are also those that have the greatest capacity to super-infect (i.e., if σ increases with α, see Sect. 6.4.1). High transmission can then be guaranteed either by being very contagious to susceptible individuals or by being very efficient at contaminating hosts that are already infected (Nowak and May 1994). In this situation as in the case we detail above, strain with a R_0 below one can be maintained in polymorphic populations of pathogens.

6.7 Concluding Remarks

In this chapter we study the evolution of parasite virulence under different scenarios. We show that the basic reproductive ratio, R_0, is a valid measure of pathogen fitness only under very restrictive conditions (e.g., no superinfection, no vertical transmission). Hence the intuitive, and widespread, reasoning based on the idea that

evolution increases R_0 often yields wrong predictions. This reasoning stems from a confusion between epidemiological and evolutionary dynamics. The basic reproductive ratio R_0 is an epidemiological quantity. To analyze evolutionary dynamics and derive a measure of pathogen's fitness it is best to go back to classical population genetics framework. We briefly present this alternative approach. This allows to rephrase the evolutionary dynamics in the classical quantitative framework where the evolution of the pathogen depends on the gradient of selection and the G matrix (i.e., genetic variance covariance matrix) of pathogens traits. In principle, the gradient of selection could be derived from a careful description of the life cycle of the pathogen. To make short term predictions of the evolutionary dynamics of pathogens we would only need to quantify G.

In the past few years, molecular biologists have studied the details of molecular interactions between pathogens and their hosts. Their complex descriptions of regulatory networks is a way to understand the proximal mechanisms that determine virulence, transmission and recovery in pathogens. They can also help to understand why and how these traits covary. This is precisely this covariation that the matrix G tries to describe and quantify. We would like to advocate here that these models will gain considerable predictive power when we will be able to combine these two approaches. Theoretician indeed still lack data to refine their description on the constraints that relate virulence, transmission and recovery. We recognize that quantifying the amount of genetic variation (and covariation) in natural populations of pathogens is a daunting task. But quantifying G is probably feasible in controlled evolution experiment studies, using model systems which molecular biology is already well understood.

References

Alizon S, van Baalen M (2005) Emergence of a convex trade-off between transmission and virulence. Am Nat 165:e155–e167

Alizon S, van Baalen M (2008) Transmission-virulence trade-offs in vector-borne diseases. Theor Popul Biol 74:6–15

Alizon S, Hurford A, Mideo N, van Baalen M (2008) Virulence evolution and the trade-off hypothesis: history, current state of affairs and future. J Evol Biol 22:245–259

Andre JB, Ferdy JB, Godelle B (2003) Within-host parasite dynamics, emerging trade-off, and evolution of virulence with immune system. Evol 57:1489–1497

Day T, Gandon S (2006) Insights from price's equation into evolutionary epidemiology. In: Feng Z, Dieckmann U, Levin S (eds) Disease evolution: models, concepts, and data analyses. American Mathematical Society, Providence, pp 23–44

Day T, Gandon S (2007) Applying population-genetic models in theoretical evolutionary epidemiology. Ecology Letters 10:876–888

Day T, Proulx S (2004) A general theory for the evolutionary dynamics of virulence. Am Nat 163:E40–E63

Day T, Graham A, Read A (2007) Evolution of parasite virulence when host responses cause disease. Proc R Soc B-Biol Sci 274:2685–2692

de Roode J, Culleton R, Cheesman Sa (2004) Host heterogeneity is a determinant of competitive exclusion or coexistence in genetically diverse malaria infections. Proc R Soc Lond Ser B-Biol Sci 271:1073–1080

Dieckmann U, Metz J, Sabelis M, Sigmund K (eds) (2002) Adaptive dynamics of infectious diseases: in pursuit of virulence management. Cambridge studies in adaptive dynamics. Cambridge University Press, Cambridge, MA
Feng Z, Dieckmann U, Levin S (eds) (2006) Disease evolution: models, concepts, and data analyses. American Mathematical Society, Providence
Ferdy JB, Godelle B (2005) Diversification of transmission modes and the evolution of mutualism. Am Nat 166:613–627
Frank SA (1996) Models of parasite virulence. Q Rev Biol 71:37–78
Gandon S (2004) Evolution of multihost parasites. Evol 58:455–469
Gandon S, Jansen VAA, van Baalen M (2001) Host life-history an the evolution of parasite virulence. Evol 55:1056–1062
Ganusov V, Antia R (2003) Trade-offs and the evolution of virulence of microparasites: do details matter? Theor Popul Biol 64:211–220
Kamo M, Sasaki A, Boots M (2007) The role of trade-off shapes in the evolution of parasites in spatial host populations: an approximate analytical approach. J Theor Biol 244:588–596
Lande R (1976) Natural-selection and random genetic drift in phenotypic evolution. Evol 30:314–334
Lande R, Arnold S (1983) The measurement of selection on correlated characters. Evol 37:1210–1226
Levin SA, Pimentel D (1981) Selection of intermediates rates of increase in parasite-host systems. Am Nat 117:308–315
Lipsitch M, Nowak M, Ebert D, May R (1995) The population-dynamics of vertically and horizontally transmitted parasites. Proc R Soc Lond Ser B-Biol Sci 260:321–327
Lipsitch M, Siller S, Nowak MA (1996) The evolution of virulence in pathogens with vertical and horizontal transmission. Evol 50:1729–1741
Nowak M, May R (1994) Superinfection and the evolution of parasite virulence. Proc R Soc Lond Ser B: Biol Sci 255:81–89
Price G (1970) Selection and covariance. Nature 277:520–521
Roff D (1992) The evolution of life histories: theory and analysis. Chapman and Hall, New York
Stearns S (1992) The evolution of life histories. Oxford University Press, Oxford
van Baalen M, Sabelis M (1995) The dynamics of multiple infection and the evolution of virulence. Am Nat 146:881–910

Part IV
Case Studies

Chapter 7
Molecular Epidemiology of Tuberculosis

Philippe Lanotte

Abstract Tuberculosis remains a major public health problem. The pathology caused by a bacteria belonging to the Mycobacterium tuberculosis complex is responsible of more than nine million new cases and of nearly two million deaths per year. The development of molecular biology since early 1990s and recent advances according to genomics offer new opportunities to understand the epidemiological dissemination of strains from patient scale to the world scale. Molecular methods were initially developed to confirm genetic link between M. tuberculosis strains isolated in similar epidemiological circumstances such as intra familial transmission, nosocomial transmission, distinction between exogenous re-infection or relapse and to explore suspected transmission chain. Methods were first based on analysis of polymorphism of an insertion sequence IS6110 by southern blotting, which evolved to be the gold standard for genotyping. PCR-based methods were developed mainly with IS6110 as target. A method based on the analysis of the Direct Repeat (DR) region, further named spoligotyping, allows identification and typing of M. tuberculosis complex isolates at the same time. In years 2000, exploration of Variable Number of Tandem Repeat (VNTR), called MIRU for Mycobacterial Interspersed Repeat Unit, was developed. This method consists in amplifying polymorphic repetitive sequences scattered throughout M. tuberculosis chromosome by PCR, in order to obtain a digit corresponding to the repetitions present at each locus. For phylogenetic purposes, all these molecular methods based on mobile genetic elements, especially insertion sequences, or repetitive DNA sequences showed limits and were supplanted by Single Nucleotide Polymorphisms (SNP), Large Sequence Polymorphism (LSP) also called Regions of Differences (RD).

P. Lanotte (✉)
Service de Bactériologie-Virologie, Hôpital Bretonneau, CHRU de Tours,
2 boulevard Tonnellé, Tours Cedex 37044, France

Laboratoire de Microbiologie, Faculté de Pharmacie & EA 3854, Faculté de Médecine,
Université François Rabelais, Tours, France
e-mail: philippe.lanotte@univ-tours.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_7,

7.1 Introduction

Tuberculosis is a scourge shared by humankind since at least thirty-five thousand years (Gutierrez et al. 2005; Daniel 2006). This pathology exploded especially in Europe in nineteenth–twentieth centuries and a quarter of Europeans died from tuberculosis at this period (Daniel 2006). With the improvement of hygiene, prevention by BCG and anti-tuberculosis therapies in sixties, it was guessed than this disease would disappear by 2000. This objective was not reached due to global poverty and apparition of HIV/AIDS pandemic.

Actually, tuberculosis is still a major cause of illness and death worldwide, especially in Asia and Africa. Globally, 9.2 million new cases and 1.7 million deaths from tuberculosis occurred in 2006, of which 0.7 million cases and 0.2 million deaths concerned HIV-positive people. Population growth has boosted these numbers compared with those previously reported by the World Health Organization (WHO report 2009). India, China, Indonesia, South Africa and Nigeria rank in top countries in terms of absolute numbers of cases. The African region has the highest incidence rate (363 per 100,000 population, Fig. 7.1). There was an estimation of 14.4 million prevalent cases of tuberculosis in 2006, with 0.5 million cases of multidrug-resistant (MDR) tuberculosis, 1.5 million deaths from tuberculosis in HIV-negative people and 0.2 million among people infected with HIV. Africa, South-East Asia and Western Pacific regions accounted for 83% of total notification cases. Several new public health problems emerged, such as VIH (especially in Africa) with 80% people co-infected tuberculosis-VIH in some countries, and the extension of MDR (multidrug resistant) and XDR (ultra-drug resistant). 440,000 cases of MDR tuberculosis were estimated, which corresponds to 3.6% of tuberculosis cases. Almost 50% of these MDR cases occurred in China and India. Moreover MDR-tuberculosis caused 150,000 deaths in 2008. Among MDR tuberculosis, 5.4% were considered XDR tuberculosis.

Mycobacterium tuberculosis is the main representative member of the Mycobacterium tuberculosis complex (MTBC). MTBC includes M. tuberculosis, M. africanum which are pathogens restricted to humans and Mycobacterium bovis has a wide host spectrum including humans. Mycobacterium canettii is also pathogen for human but is exceptionally isolated, Mycobacterium microti is known as infecting rodents. MTBC members are characterized by 99.9% similarity at the nucleotide level and identical for their 16S rRNA sequences (Boddinghaus et al. 1990; Sreevatsan et al. 1997) but they differ widely in terms of their host tropisms, phenotypes, and pathogenicity.

In recent years, with the increasing knowledge on M. tuberculosis genome, molecular methods were developed for different objectives: molecular typing using genetic markers to explore chain of transmission, to distinguish recent exogenous from reactivation of latent tuberculosis, to study MDR outbreaks, to identify species among MTBC, to conduct phylogenetic studies. The choice of markers is then essential given the limitations of certain methods and their accessibility and applicability.

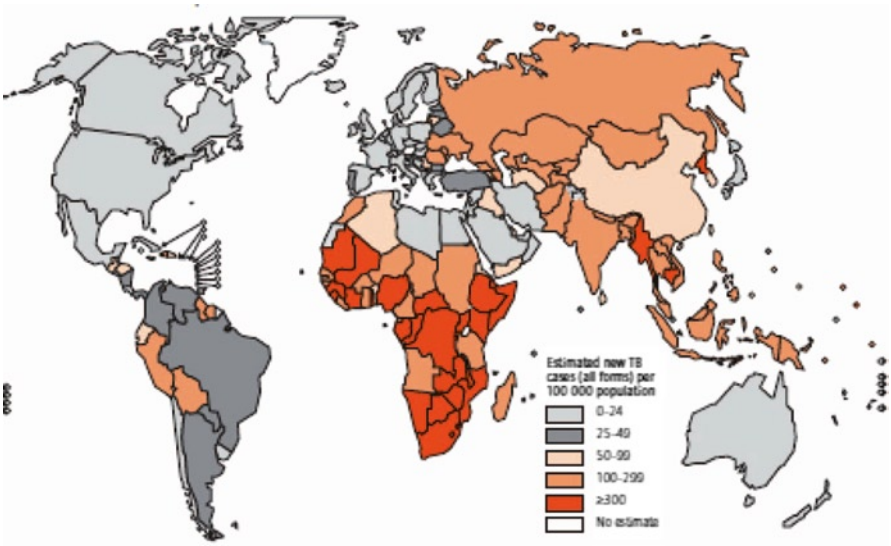

Fig. 7.1 Estimated tuberculosis incidence rates, 2008 (WHO report 2009)

7.2 Genetic Markers and Molecular Methods in the Field of Tuberculosis

7.2.1 IS6110 and Restriction Length Fragment Polymorphism (RFLP)

Among the genetic markers used, insertion sequence (IS) and specifically IS6110 was the first marker used in a typing objective (Thierry et al. 1990a; Thierry et al. 1990b; Coros et al. 2008). IS6110 is member of the family IS3. Four to 20 copies of this 1,361 bp element have been found scattered throughout the genome of M. tuberculosis, while strains of M. bovis contain only few copies of IS6110 (Cousins et al. 1998). With the data obtained from the sequencing of M. tuberculosis H37Rv, most of the IS6110 copies were found in intergenic or non-coding region, near tRNA genes. The distribution of IS6110 is not homogeneous on the chromosome (Cole et al. 1998). Initial studies confirmed that, even if IS6110 is a transposable element, transposition frequency is low and no polymorphism was detected in experimental infection of guinea pigs with M. tuberculosis strains over a period of 3 months. The stability is probably maintained for many years. The number of copies of IS6110 is variable for an isolate to another with variable localisation on the chromosome. Due to this second polymorphism level, the distribution of IS6110 in M. tuberculosis strains isolated from different patients can reveal

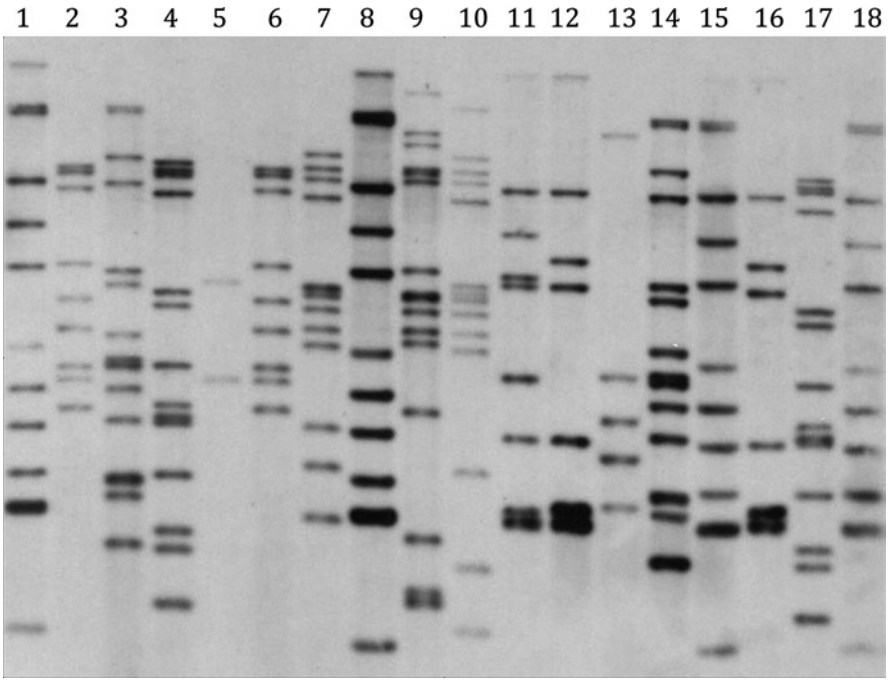

Fig. 7.2 Patterns of M. tuberculosis isolates based on the IS6110-RFLP method described by van Embden et al. 1993. *Tracks 1*, *8*, *18* correspond to M. tuberculosis 14323. *Others tracks* correspond to various isolates

different genomic arrangements of the sequences, whereas strains isolated from patients of the same tuberculosis outbreak may show identical distribution patterns (van Embden et al. 1993).

The method for the analysis of the IS6110 was named IS6110 Restriction Fragment Length Polymorphism (IS6110-RFLP) and became the reference method for typing M. tuberculosis. IS6110-RFLP profiling is performed by an internationally standardized method with restriction endonuclease PvuII, which has only one restriction site in IS6110 (van Embden et al. 1993). Fragments are separated by agarose gel electrophoresis. After a transfer on membrane, IS6110 used as a probe can hybridize with the restriction fragment length containing the IS. Hybridizing bands are visualized by enhanced chemiluminescence (Fig. 7.2). This manual method requires few milligrams of DNA and thus it is practicable only after abundant culture, and also time consuming as it needs 5 days from extraction to revelation. Patterns are stable for a same isolate because of low transposition range of IS6110, as the transposition rate is linked to the exposure to a microaerobic environment (de Boer et al. 2002; Ghanekar et al. 1999). This may explain some minor differences observed between strains of a same patient with a relapse of tuberculosis or in secondary cases of the same chain of transmission. This method is not powerful for isolates with low number of IS6110 copies.

A great number of targets that presented polymorphisms were also proposed as typing method by RFLP for tuberculosis. Nevertheless these techniques have proved less success. Targets are IS1081 (van Soolingen et al. 1992), gene mtp40 (Del Portillo et al. 1991), mpb64 (Li et al. 1993), katG (Zhang et al. 1992; van Soolingen et al. 1995) and 16S rRNA(van Soolingen et al. 1995).

7.2.2 IS6110 and Polymerase Chain Reaction Methods

To associate performance of IS6110 detection and to overcome the disadvantages of the IS6110-RFLP method, some methods using PCR targeting IS6110 were developed.

The Mixed-Linker PCR (ML-PCR) is based on the amplification of restriction length fragment containing IS6110 (Haas et al. 1993). Bacterial DNA is initially restricted by HhaI and restriction fragment are ligated with a peculiar linker at each extremity. This linker is composed of a double-strand oligonucleotide, with one strand containing uracile instead of thymidine. The uracile strand is then eliminated under the action of uracile-N-Glycosylase. The next step consists of a PCR with one primer targeting IS6110 and one primer targeting the linker. After 30 amplification cycles, a hemi-nested PCR is performed using a complementary primer of the IS6110 introduced in the first PCR and a second specific primer IS6110. Amplification patterns are compared after gel electrophoresis. This method, expensive and time consuming, possesses a discriminatory power similar to IS6110-RFLP.

Another method, also based on a ligation step was developed, and quite easy to perform. In the ligation-mediated PCR method (LM-PCR), DNA is restricted with SalI, a double-strand asymmetrical linker is ligated to the cohesive extremities for restriction sites. Then a PCR is performed with a primer targeting the linker and the other one targeting IS6110, allowing the amplification of the flanking region of IS6110 (Prod'hom et al. 1997). Patterns are then compared after gel electrophoresis (Fig. 7.3). Nevertheless, these methods do not replace the IS6110-RFLP gold standard method (Kremer et al. 1999).

7.2.3 Direct Repeat Region – CRISPR and Spoligotyping

Another method less DNA consuming than IS6110-RFLP was developed targeting a direct repeat (DR) locus (Kamerbeek et al. 1997). This method is able to detect M. tuberculosis in clinical specimens or from culture, and allow the differentiation of M. bovis from M. tuberculosis and even some sub-clades within M. tuberculosis. The DR locus contains multiple, well-conserved 36 bp DRs interspersed with non-repetitive spacer sequences of variable sizes, from 35 to 41 bp. Isolates vary in number of DRs and in the presence or absence of specific spacers (Fig. 7.4a). The chromosome of M. tuberculosis H37Rv contains 48 conserved DRs. By comparison,

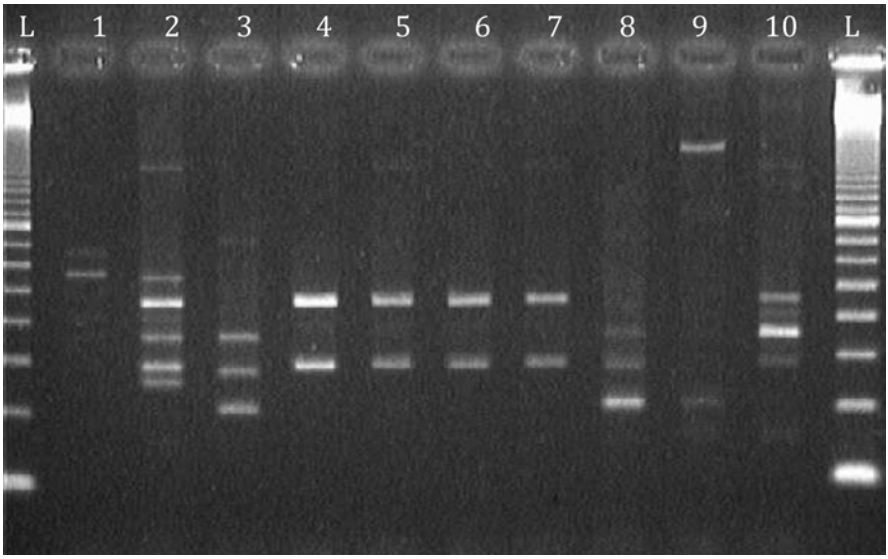

Fig. 7.3 Example of result obtained by LM-PCR. *L*: DNA ladder (100 bp). *Tracks 3* and *8* correspond to related isolates in a case of health worker contamination. *Tracks 5*, *6* and *7* correspond to laboratory contamination. The sample source of the contamination presented a positive direct examination over 100 FAB/microscopic field. The corresponding strain pattern is on *track 4*

M. bovis BCG contains 41 DRs. The DR locus is present only in bacteria belonging to MTBC. Extensive analysis of DR locus allows identification of 43 intergenic sequences. The method is facilitated by the commercialisation of membranes containing the 43 oligonucleotides sequences of the spacers. The method is referred to as spacer oligotyping or "spoligotyping", because it is based on isolate-dependent hybridisation patterns of amplified DNA with spacer oligonucleotides. Moreover, spoligotyping is more efficient for M. tuberculosis isolates with low IS6110 copies. By example among 19 M. tuberculosis isolates with only one IS6110 copy, ten spoligotypes were detected. Nevertheless, IS6110-RFLP is able to better discriminate isolates containing more than five IS6110 copies (Kamerbeek et al. 1997; Diaz et al. 1998).

Thus, DNA is extracted from clinical samples or from a MTBC positive culture, and amplified with a couple of primer including one's biotinylated. These primers target any DRs in the DR region and the amplified DNA is composed of a mixture of a large number of different-size fragments including the interspersed spacers. The biotinylated amplified DNAs are then hybridised directly on the membrane and are covalently bound to oligonucleotides when the spacers were present in the isolate DNA. The spacers are ordered in the membrane in the same was as there are ordered in the genome. The presence of a spacer in the M. tuberculosis DNA tested was highlighted when the signal is dark (Fig. 7.4b). The spacer is considered absent in case of absence of hybridisation. The detection is easy, with a binary code, presence or absence for each of the 43 spacers (Kamerbeek et al. 1997). Thus, isolates from

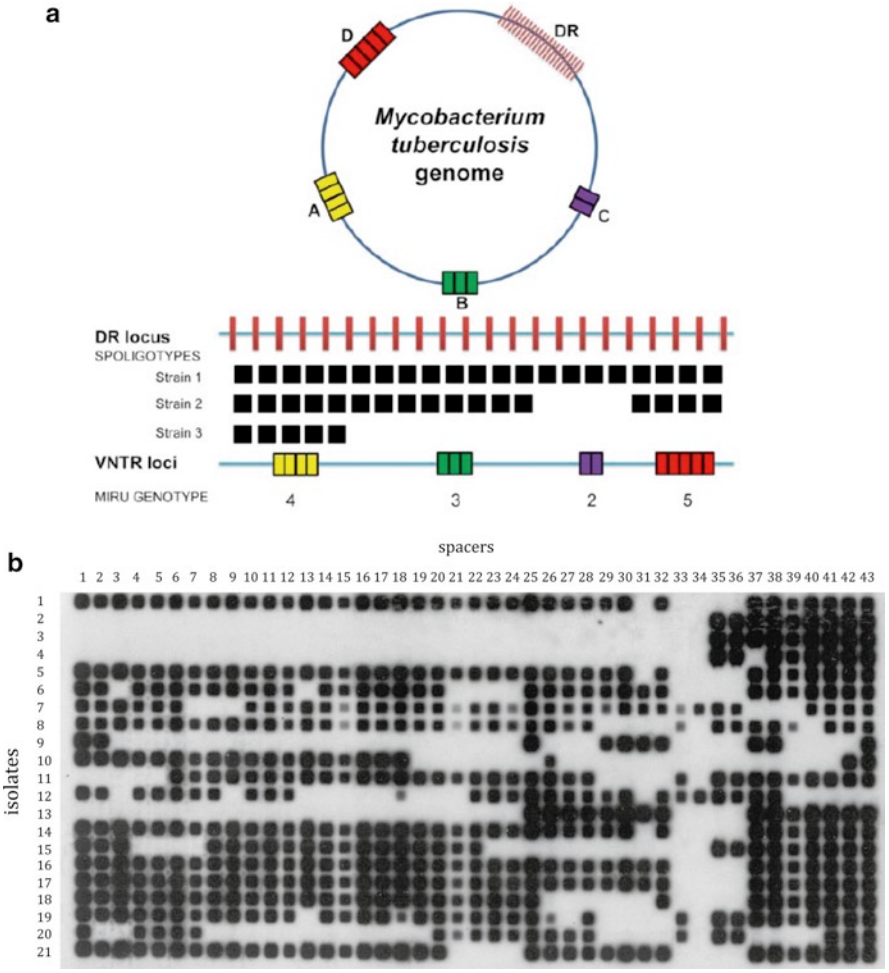

Fig. 7.4 (**a**) A-Schematic illustrating the principles of the CRISPR- and VNTR-based genotyping in MTBC. These genotyping methods are known as 'spoligotyping' and 'MIRU-VNTR-typing', respectively. Spoligotyping is based on the detection of 43 unique spacers located between direct repeats at a specific locus of the MTBC genome known as the direct repeat (DR) locus. Spoligotyping patterns are commonly represented by *black* and *white squares* indicating presence or absence of particular spacers, respectively. The deletion of some of these 43 spacers allows to differentiate between strains. MIRU-VNTR analysis relies on the identification of different number of repeats at several loci scattered around the bacterial genome (marked by (**A**), (**B**), (**C**), and (**D**) in the figure). The number of repeats at each locus is combined to generate a unique numerical code used to establish phylogenetic and epidemiological links between strains (from Comas et al. 2009). (**b**) Example of spoligotype patterns

epidemiologically-related cases share same hybridisation patterns. An international spoligotyping database (SpolDB4) has been recently made accessible via internet, which facilitates molecular epidemiological studies at a global scale (Brudey et al. 2006).

The Direct Repeat locus of the MTBC is a member of the CRISPR (Clustered regularly interspaced short palindromic repeats) sequences family recently described (Brudey et al. 2006; Pourcel et al. 2005). Spoligotyping is a widely used PCR-based reverse-hybridization blotting technique that assays the genetic diversity of this locus and useful for clinical laboratory, molecular epidemiology, evolutionary and population genetics. Recently a new format of this method has been proposed as a test that uses microbeads-based techniques with new spacers to increase the discriminative power of the method (Zhang et al. 2010).

7.2.4 Short Repetitive DNA Sequence and Polymorphic GC-Rich Sequence

Short repetitive DNA sequences associated with genetic diversity were also used as typing methods. Repeat of a triplet GTG was detected by southern blotting after HinfI digestion (Wiid et al. 1994). Polymorphic GC-rich repeat sequence (PGRS) also called Pro-Glu (Prolin-Glutamic acid or PE) repeat sequence is the target of southern-blotting methods after AluI digestion (Ross et al. 1992). In M. tuberculosis, PGRS appears to be present in at least 30 copies varying in number and distribution from strain to strain.

Double Repetitive Element amplification (DRE-PCR) targets PGRS and IS6110 (Friedman et al. 1995). This technique is based on the fact that the distance between the copy numbers of IS6110 and PGRS may vary from strain to strain. These variations allow different sizes and numbers of DNA fragments to be amplified.

7.2.5 Analysis of Tandem Repeats, VNTR-MIRU Method

Major Polymorphic Tandem Repeat (MPTR), present at multiple chromosomal loci, was used initially to compare isolates (Hermans et al. 1992). Eleven loci of tandem repeats were then explored by Frothingham et al. (Frothingham et al. 1998). Among them, six Exact Tandem Repeats (ETR), ETR-A to ETR-F and one among five MPTR (MPTR-A) present a sufficient polymorphism. All together, these markers were able to distinguish 22 of 25 strains of M. tuberculosis and five on 23 M. bovis BCG tested. Each ETR locus had multiple alleles in the panel. Polymorphism corresponds to insertions or deletions of tandem repeats. Allele profiles were reproducible and stable, as demonstrated by analyses of multiple isolates of particular reference strains (Frothingham et al. 1998). Supply et al. described 41 Variable Number of Tandem Repeat (VNTR) called MIRU (Mycobacterial Interspersed

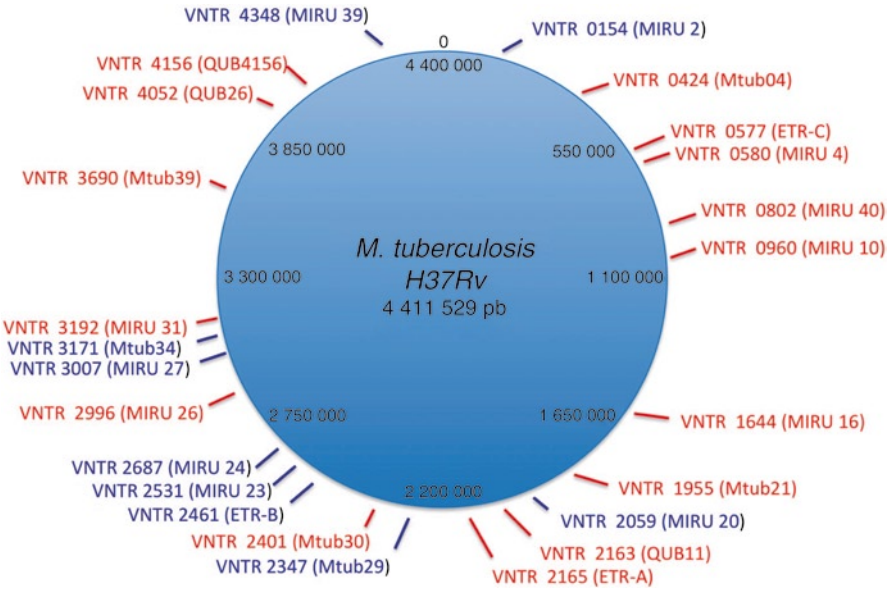

Fig. 7.5 Position of the 24 loci set proposed as standardisation for MIRU-VNTR typing (Supply et al. 2006). The 15 MIRU-VNTR loci proposed for epidemiological discrimination are in *red* and of the nine supplementary MIRU-VNTR loci for a higher resolution are in *blue*. They are named by their position on H37Rv chromosome are synonymous names are in *brackets* (Modified from Supply et al. 2000)

Repeat Unit) (Supply et al. 1997; Supply et al. 2000; Mazars et al. 2001). These MIRU corresponding to minisatellite-like structures are composed of 40–100 bp repetitive sequences that are scattered throughout the chromosome of M. tuberculosis H37Rv strain (Fig. 7.5). Two of the 41 loci, VNTR0580 and VNTR 3192 correspond to ETR-D and ETR-E respectively. Among them, 12 loci contain variable copy numbers of 51-to-77 bp MIRUs that were selected for genotyping (Supply et al. 1997; Supply et al. 2000; Mazars et al. 2001). Globally, the method combining ETR, MPTR and VNTR is called MIRU-VNTR. Each locus is amplified by PCR (Fig. 7.6). Results are reported in a 12-digit format corresponding to the number of repeats at each chromosomal locus. This data format allows inter- and intra-laboratory comparisons. Nevertheless, various nomenclature from different laboratories create some confusion as some TR are identical but with different names (Le Flèche et al. 2002; Smittipat et al. 2005; Supply et al. 2006).

The number of allele for each VNTR loci may vary from 1 or 2 (MIRU 2 or VNTR0154) up to 32, for VNTR 3820 (Smittipat et al. 2005; Supply et al. 2006; Sun et al. 2004). Among all MIRU-VNTR described, many scheme were proposed for genetic comparison of isolates (Supply et al. 2006). The polymorphism of each locus may vary within different isolates due to the geographical origin, or in peculiar population. Nevertheless, to be comparable, a 15-locus system was proposed for epidemiological study. This number of markers was considered enough discriminative

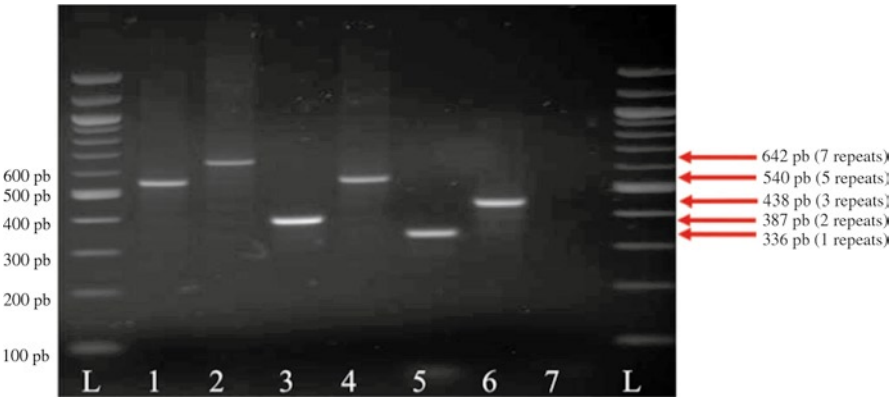

Fig. 7.6 Example of patterns obtained for MIRU 26. *L*: DNA Ladder (100 pb). *Tracks 1–5*: clinical isolates, *track 6*: M. tuberculosis H37Rv, *track 7*: negative control

for routine investigations, but a 24-locus system was considered as a high-resolution tool for phylogenetic studies (Supply et al. 2006).

MIRU-VNTR is easy to perform and accessible method for genetic studies (Martin et al. 2007). The initial method can be performed on automated sequencer with a peculiar tool to identify the PCR product size when fluorescent dye primers are used. Moreover, using automated sequencer, three or four VNTR can be explored simultaneously (Supply et al. 2001).

7.2.6 Single Nucleotide Polymorphisms, the SNP Method

Genetic polymorphism can be detected also at the nucleotide level. Single nucleotide polymorphisms (SNP) are decomposed in non synonymous (nsSNP) and in synonymous SNPs (sSNP). nsSNP indicate amino acid changes and are implicated by example, in resistance to anti-tuberculosis drugs (Gutacker et al. 2006). These sSNP are able to divide mycobacterial populations in genetic groups. Moreover, sSNP are theoretically evolutionarily neutral and then can be used for population-genetics and for examination of phylogenetic relationships among bacterial isolates. Combination of sSNP, nsSNP and intergenic SNP was shown to be representative of phylogenetic lineages over the world (Gutacker et al. 2006).

Phylogenetic studies undergone with this method seem to be robust (Comas et al. 2009). Among 212 SNPs tested, Filliol et al. (2006) identified six SNP markers that were sufficient to classify a global M. tuberculosis collection into seven phylogenetically distinct "SNP cluster groups" (SCGs). Using some others SNPs enabled to subdivide the SCGs into subgroups (SC subgroups) giving then a more discriminative method for phylogenetic analysis (Alland et al. 2007).

Recently, an approach combining two multiplex allele-specific minisequencing assays was proposed and permits detection of eight species- and eight lineage-specific

single nucleotide polymorphisms (SNP) (Bouakaze et al. 2010). This method uses a commercial kit SNaPshot™ Multiplex Kit (Applied Biosystems) followed by an eight-plex minisequencing reaction and finally analysis of extension products by capillary electrophoresis. Distinction between principal genetic groups defined by SNPs was then obtained for 56 MTBC strains (Bouakaze et al. 2010).

7.2.7 Large Sequence Polymorphisms or Regions of Difference, Comparative Genomic and DNA Microarrays

Comparative genomics studies, available since the first publication of a complete genome of M. tuberculosis H37Rv strain (Coros et al. 2008), revealed that large genomic insertions or deletions are important sources of genome plasticity in MTBC (Fleischmann et al. 2002; Brosch et al. 2002). These genomic modifications are also known as large sequence polymorphisms (LSPs) or regions of difference (RD) (Brosch et al. 2002; Tsolaki et al. 2004). More than 60 different LSPs were found in clinical isolates in comparison with M. tuberculosis H37Rv representing around 4% of the genome (Tsolaki et al. 2004). Deletions are not randomly distributed. Some LSPs are associated with genes encoding members of the proline-glutamic acid or proline-proline-glutamic acid protein families and represent polymorphism that occurs under selective pressure while others may reflect phylogeny (Alland et al. 2007). RD were also used for identification of species within MTBC (Brosch et al. 2002; Huard et al. 2006; Pinsky and Banaei 2008). DNA microarrays were used to analyze these LSPs and to improve evolutionary knowledge of M. tuberculosis (Comas et al. 2009; Tsolaki et al. 2004; Hirsh et al. 2004; Gagneux et al. 2006).

7.3 Applications of Molecular Typing for Tuberculosis

Applications of molecular typing techniques are various from inter-human transmission with several objectives (MDR transmission, intra-familial transmission, re-infection/relapse, nosocomial transmission) to evolution of M. tuberculosis strains at global scale, including laboratory contamination and patho-physiology among others. The main applications of these methods are discussed below.

7.3.1 Inter-Human Transmission

Initial genetic studies were performed to confirm outbreaks, to identify index cases in outbreak, to trace chain of transmission in specific groups (like prisoners), and to identify unsuspected transmission. Studies have explored epidemiology within a town or at the level of a country (Diaz et al. 1998; Small et al. 1994;

Lopez-Calleja et al. 2009). IS6110-RFLP, Spoligotyping and MIRU-VNTR were the predominantly methods used for this purpose. Molecular methods also showed that evolution of M. tuberculosis in vivo is characterized by periods of relative genomic stability followed by bursts of mutation (Schurch et al. 2010). Molecular epidemiology improved knowledge on the dissemination of the bacteria. For example, in Germany, Barniol et al. (2009) demonstrated that few exchanges of strains between foreign-born cases in comparison to exchanges within local populations (Barniol et al. 2009).

7.3.2 Multi-Drug Resistant Transmission

Many studies have focused on the transmission of MDR strains. In 1993, a major outbreak was observed in New York and 60% of the cases were linked to MDR strains (Valway et al. 1994; Frieden et al. 1996). Cases diagnosed accounted for nearly one fourth of the cases of multidrug-resistant tuberculosis in the United States over a 3 years period. Most patients had nosocomial-acquired disease, and were also infected by HIV (86%), with an important mortality rate (83%). IS6110-RFLP typing method used to characterized this peculiar strain ("strain W") and some variants allowed the follow-up of the dissemination of the strain widely in the community over years (Munsiff et al. 2003) across the United States (Agerton et al. 1999) and was also isolated in Europe (Bifani et al. 1996; Schwoebel et al. 1998). Variants observed on molecular patterns presented additional drug resistance probably linked to nsSNP as demonstrated by other studies (Bifani et al. 2008; Post et al. 2004). Moreover, new insights concerning local phenomenon in lung lesions lead to the emergence of heterogeneous population of bacilli with different drug-susceptibilities (Post et al. 2004; Kaplan et al. 2003).

A well-known worldwide cluster associated with anti-tuberculosis drug resistance is the Beijing genotype, which can be identified routinely by spoligotyping. This peculiar genotype, genetically linked with the "strain W", is mainly responsible of MDR and XDR strains over the world. For example, in a recent study in Taiwan, 44.4% of the XDR M. tuberculosis isolates and 56.7% of the MDR isolates belonged to the Beijing genotypes' family (Lai et al. 2010).

7.3.3 Exogenous Re-infection and Relapse

It was admitted that initial tuberculosis protects against re-infection. Molecular methods were able to demonstrate that re-infection exists and that the relative importance of re-infection is likely depending on the epidemiological context. In geographical areas with a low incidence of tuberculosis, recurrent tuberculosis is generally due to reactivation of the disease. An increased risk for re-infection was

observed in immigrant patients compared to inhabitants in low tuberculosis incidence areas (Mathema et al. 2006; Warren et al. 2004).

A molecular epidemiology study performed in Italy, based on more than 2,100 patients with tuberculosis, identified that 32 patients (1.5%) had two distinct episodes of tuberculosis (with a cure as outcome of the first episode and 6 months between the two episodes) (Bandera et al. 2001). In five patients (16%), the DNA fingerprinting patterns of M. tuberculosis strains responsible of the second episode did not match those corresponding to isolates of the first episode, indicating exogenous re-infection (Bandera et al. 2001). Episodes of re-infection in areas with low incidences of tuberculosis are, however, rare compared to those in high-incidence geographical regions. It was also admitted that tuberculosis may result from a single infection with a single M. tuberculosis strain. A study suggested that multiple infections are frequent in high-incidence regions, implying high re-infection rates and lack of efficient protective immunity conferred by initial infection (Warren et al. 2004). In populations that have emigrated from high-risk areas, re-infection may represent a considerable contributor to the rate of recurrent tuberculosis. Moreover, drug susceptible and drug-resistant strains may also coexist (Warren et al. 2004).

Relapse is not an anecdotal aspect of tuberculosis. Between 1% and 11% of cases have a second recurrent episode (Martin et al. 2007). Relapse could be associated with non observance of therapy, immune-suppressive therapy or infection in the elderly (Hocking and Choi 1997; Mathema et al. 2006; Comas and Gagneux 2009).

Identification of recurrences caused by exogenous re-infection could influence therapeutic and epidemiological decisions because susceptibility could be different and the patient should be considered as a new case. When a case is assumed to be a relapse, rapid information on exogenous re-infection by a strain spreading into the community could indicate new recent transmission routes and ongoing transmission events. Moreover, higher risk of relapse rather than re-infection may be observed in HIV-positive subjects and in patients infected with multidrug-resistant tuberculosis.

M. tuberculosis isolates exhibiting identical DNA fingerprinting patterns can harbor substantial genomic diversity. Because this variability may not be captured by traditional genotyping methods of MTBC, some important aspects of the transmission dynamics could be missed or misinterpreted (Niemann et al. 2009).

7.3.4 Nosocomial Transmission

During the New York's outbreak of 1990–1993, the great majority of cases were acquired at hospital (Valway et al. 1994; Frieden et al. 1996). Dissemination of the "W strain" in other states was associated with the use of bronchoscope (Agerton et al. 1999). In France, M. tuberculosis cross-contamination due to the use of bronchoscope was also confirmed by molecular technique (Carricajo et al. 1999).

The risk of transmission of M. tuberculosis from patients to health-care workers is a neglected problem in many countries of low- and middle incomes especially when the clinical form is not typical (Joshi et al. 2006; Menzies et al. 2007). A higher risk of acquiring TB disease was associated with certain work locations (inpatient TB facility, laboratory, internal medicine, and emergency facilities) and occupational categories (radiology technicians, patient attendants, nurses, ward attendants, paramedics, and clinical officers) (Joshi et al. 2006).

In high-income countries, risk can be higher for health-care workers if the infection control measures are ineffective (Menzies et al. 2007).

7.3.5 Laboratory Contamination

The false diagnosis of tuberculosis due to laboratory cross-contamination is a well-known event, which has been reported to occur in 0.1–3% of cases (de Boer et al. 2002; Ruddy et al. 2002; Small et al. 1993). Laboratory contamination is suspected (i) when only one sample of a patient is culture positive with a small number of colonies, (ii) when M. tuberculosis is cultivated from a sample processed together with a smear-positive specimen, and (iii) when clinician considered an alternative diagnosis as more probable. Suspicion of false positivity is increased when these conditions are associated (Martinez et al. 2006). Microbiologists should rule out with clinicians the epidemiological links between suspected cases. The final confirmation of false positivity requires the application of molecular tools to prove that MTBC isolates from co-processed specimens share identical genotypic patterns (Martin et al. 2008). MIRU-VNTR appears to be more adequate than RFLP for analyzing cross-contamination alerts. It is faster than RFLP and the correlation with RFLP diagnosis is high. A permanent suspicious attitude from clinical bacteriologist and an access to fast resolution of cross-contamination alerts could enable more rapid management of suspected false positive cases (Mathema et al. 2006).

7.3.6 Phylogeny of M. tuberculosis at Global Scale

As mentioned above, molecular tools and appropriate uses of new markers allow the determination of the emergence of peculiar clones over the world.

Although strain W was characterized in New York City using IS6110-RFLP (Agerton et al. 1999), a similar group of predominant isolates was identified with the same molecular characteristics in China. IS6110-RFLP showed that 86% of the isolates belonged to a genetically closely-related group (van Soolingen et al. 1995; Bifani et al. 2002). Because the majority of these strains originated from the province of Beijing, the cluster was named the “Beijing family”. Strains of this family were also found to dominate in neighboring countries such as Mongolia, South Korea, and Thailand, where this cluster represents respectively 50%, 43%, 37%

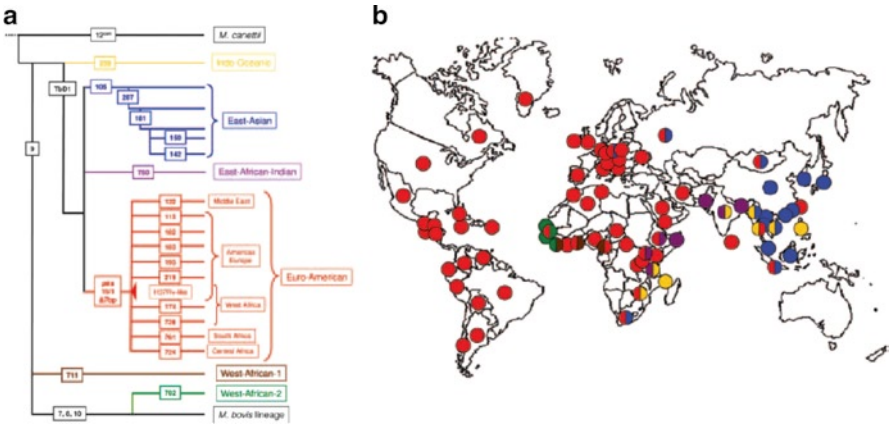

Fig. 7.7 The global population structure and geographical distribution of M. tuberculosis. (**a**) LSPs define a global phylogeny for M. tuberculosis. The names of the lineage-defining LSPs or regions of difference are shown in rectangles. The geographic regions associated with specific lineages are indicated. (**b**) The six main lineages of M. tuberculosis are geographically structured. Each *dot* corresponds to 1 of 80 countries represented in the global strain collection. *The colors of the dots* relate to the six main lineages defined in Fig. 7.7a and indicate the dominant lineage(s) in the respective countries (From Gagneux et al. 2006)

of the isolates. By contrast, low prevalence of such strains was observed in other continents (van Soolingen et al. 1995). Using spoligotyping, this cluster is easy to distinguish because of the loss of 34 spacers corresponding to deletion of RD207 (Comas et al. 2009; Tsolaki et al. 2005). Based on specific spoligotypes patterns, reports have described several strains' families such as EAI (East-African-Indian), Beijing, CAS (Central-Asian), LAM (Latin-American-Mediterranean), Haarlem, 'Cameroon', 'Uganda', 'X' among others (Filliol et al. 2002, 2003).

Recent phylogenetic analyses based on LSPs have identified six main lineages (Gagneux et al. 2006). Each of these six lineages was found in particular geographical areas (the lineage names reflect these geographical associations) (Fig. 7.7). The West-African lineage one and linage two correspond to strains named M. africanum (Comas et al. 2009; Gagneux et al. 2006). The East-Asian (or Central-Asian, CAS) lineage is dominant in many countries of East Asia and Indo-Pacific. The Euro-American lineage is clearly the most frequent lineage in Europe and the Americas, but specific sub-lineages within the Euro-American lineage predominate in different regions of Africa and the Middle East. Interestingly, the six main lineages were represented in Africa. Multi-Locus Gene sequences analysis, LSPs correlate partially with grouping obtained by spoligotyping and SNPs (Comas et al. 2009; Gagneux et al. 2006; Comas and Gagneux 2009; Hershberg et al. 2008).

The worldwide distribution of the Beijing family, and its genetic homogeneity, suggests that the strains belonging to this group might present a selective advantage. Beijing strains represent about 50% of strains in East Asia and at least 13% of strains worldwide. Their emergence might be linked to escape from BCG vaccination,

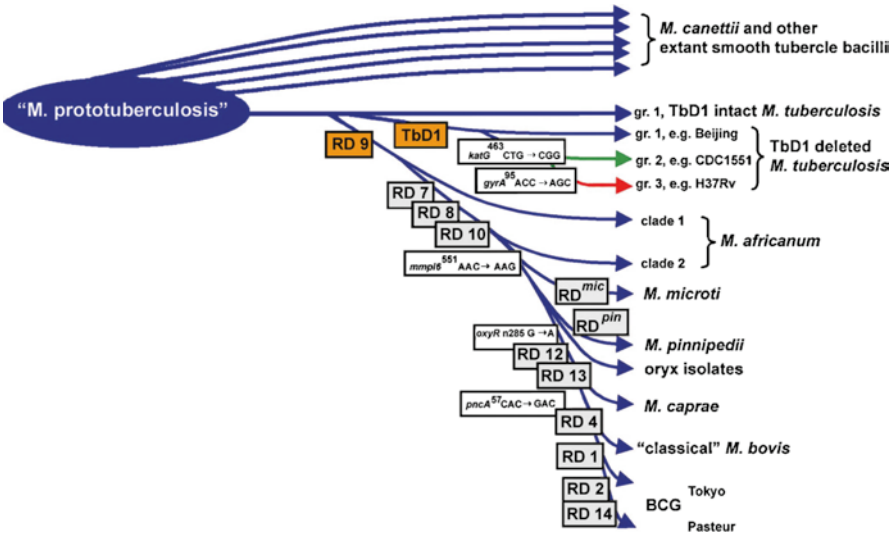

Fig. 7.8 Working model of the evolutionary scheme of tubercle bacilli illustrating successive loss of DNA in certain lineages (*grey boxes*). The scheme is based on the presence or absence of conserved RDs (TbD1) and on sequence polymorphisms in five selected genes. Note that the distances between certain branches may not correspond to actual phylogenetic differences calculated by other methods (From S. V. Gordon et al. 2009)

and to multi-drug resistance in many areas. Finally, the Beijing family has been linked to polymorphisms in immune genes, suggesting the possibility of human–mycobacterial co-evolution. These results suggest that M. tuberculosis Beijing strains have a high capacity to withstand tuberculosis treatment, even in the absence of drug-resistance (Parwati et al. 2010).

Another facet of phylogenetic studies deals with M. tuberculosis species and members of the MTBC. LSPs (RDs) helped at confirming a common ancestor, although still in debate (Brosch et al. 2002; Huard et al. 2006; Fabre et al. 2004; Gutierrez et al. 2005). Notion of ancestral (basal) and modern (derived) lineages is based on the lack of deletion in the TbD1 region containing a leucine (CTG) at katG in position 463, which characterised ancestral (basal) M. tuberculosis strains (Fig. 7.8) (Brosch et al. 2002; Gordon et al. 2009). A study has suggested that ancestral mycobacteria may have affected early hominids in East Africa around three million years ago (Gutierrez et al. 2005).

7.3.7 Pathophysiology of Tuberculosis

Molecular tools provide further insights on pathophysiology of tuberculosis. Nearly 20% of patients with a smear-negative are responsible of tuberculosis transmission (Behr et al. 1999).

Moreover, histological examination of different lung lesions revealed variability in morphology and distribution of acid-fast bacilli. Molecular characterization of different isolates from different anatomical locations (for example at the surface of cavities, in granulomas) suggests that a single infected strain may undergo genetic changes during treatment, which leads to acquisition of additional drug resistance independently in these different locations resulting in parallel evolution of heterogeneous drug-resistant sub-populations (Kaplan et al. 2003).

As mentioned previously, repetitive sequences of PE-PGRS are used in some molecular methods. PE-PGRS may have role in pathophysiology of tuberculosis. PE-PGRS16 and PE-PGRS26 can be implicated in latency phenomenon (Talarico et al. 2008). Moreover, sequence variations in the PE-PGRS33 protein with large insertion, deletion or mutation may result in the lack of cavity formation in the lungs (Talarico et al. 2007).

Some particular groups such as the Beijing/W strains may present some specific biological features. This cluster can be associated with the development of extrathoracic localization of tuberculosis (Kong et al. 2006). Moreover, the genotype of M. tuberculosis, defined by LSPs or SNPs, may influence features of pulmonary and meningitis tuberculosis. The association between Beijing lineage and disease progression suggests that this lineage may influence intra-cerebral inflammatory response. In addition, increased drug resistance among bacteria of the East Asian Beijing lineage might influence the response to treatment (Thwaites et al. 2008). Beijing strains appear to be more virulent in animal models, and to cause more histo-pathological changes, higher growth, and increased mortality. At a molecular level, Beijing strains have specific properties in terms of protein and lipid structures and interaction with the immune system (Parwati et al. 2010). All these studies suggest that the genetic diversity of M. tuberculosis has important clinical consequences.

7.3.8 Molecular Methods Adapted to Other Members of the Mycobacterium tuberculosis Complex

Molecular methods are applied for all members of MTBC. Nevertheless, some adaptations are necessary to increase the discriminatory power of markers used (Allix et al. 2006; Roring et al. 2002). Indeed, M. bovis contains few copies of IS6110 and RFLP based on this IS was found sufficiently discriminative. New loci for VNTR were identified (Roring et al. 2002); their analyses and spoligotypes patterns helped at identifying person-to-person transmission (Evans et al. 2007; Sunder et al. 2009).

In M. africanum, a new generation of spoligotyping based on a 68-spacer format defined new additional patterns, which helped at better understanding the evolution of M. africanum (Brudey et al. 2004). Complementary studies also allowed reclassification of some M. africanum lineages, initially defined as subtype (Brudey et al. 2004).

7.4 Conclusion

The standard molecular epidemiological typing techniques are based on mobile (e.g., IS6110-RFLP) or repetitive (e.g., spoligotyping and MIRU-VNTR) DNA elements that evolve fast, resulting in high discriminatory power, an important prerequisite for detecting ongoing transmission, identifying laboratory cross-contaminations, or differentiating disease relapse from re-infection.

However, because of rapid changes at these loci, identical finger printing patterns can emerge in unrelated strain lineages (homoplasy) as a result of convergent evolution, making it difficult to define deep phylogenetic structures unambiguously. Moreover, these methods could present some limits in phylogenetic studies. SNP analysis or comparative genomic studies, identifying regions of difference, seem to be more congruent and robust for this purpose.

All together, these methods allow new insights in the understanding the various facets of tuberculosis. Molecular typing methods lead to improvements in epidemiological studies in association with classical epidemiologic approaches. Genotyping methods afford greater resolution in the detection of unsuspected transmission and in the differentiation between exogenous re-infection and relapse. New genomics technologies allow the improvement of large population genetics studies and in understanding MTBC strains evolution. Identification of particular lineages or strains exhibiting specific virulence properties or epidemic potential can be described. Molecular studies have also challenged our view on the animal origin of human tuberculosis but also on the possibility of co-infection or re-infection after cure.

Acknowledgement The author thanks MC Gutierrez for providing some pictures.

References

Agerton TB, Valway SE, Blinkhorn RJ et al (1999) Spread of strain W, a highly drug-resistant strain of Mycobacterium tuberculosis, across the United States. Clin Infect Dis 29:85–92

Alland D, Lacher DW, Hazbón MH et al (2007) Role of large sequence polymorphisms (LSPs) in generating genomic diversity among clinical isolates of Mycobacterium tuberculosis and the utility of LSPs in phylogenetic analysis. J Clin Microbiol 45:39–46

Allix C, Walravens K, Saegerman C et al (2006) Evaluation of the epidemiological relevance of variable-number tandem-repeat genotyping of Mycobacterium bovis and comparison of the method with IS6110 restriction fragment length polymorphism analysis and spoligotyping. J Clin Microbiol 44:1951–1962

Bandera A, Gori A, Catozzi L et al (2001) Molecular epidemiology study of exogenous reinfection in an area with a low incidence of tuberculosis. J Clin Microbiol 39:2213–2218

Barniol J, Niemann S, Louis VR et al (2009) Transmission dynamics of pulmonary tuberculosis between autochthonous and immigrant sub-populations. BMC Infect Dis 9:197

Behr MA, Warren SA, Salamon H et al (1999) Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. Lancet 353:444–449

Bifani PJ, Plikaytis BB, Kapur V et al (1996) Origin and interstate spread of a New York City multidrug-resistant Mycobacterium tuberculosis clone family. JAMA 275:452–457

Bifani PJ, Mathema B, Kurepina NE et al (2002) Global dissemination of the Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol 10:45–52
Bifani P, Mathema B, Kurepina N et al (2008) The evolution of drug resistance in Mycobacterium tuberculosis: from a mono-rifampicin-resistant cluster into increasingly multidrug-resistant variants in an HIV-seropositive population. J Infect Dis 198:90–94
Boddinghaus B, Rogall T, Flohr T et al (1990) Detection and identification of mycobacteria by amplification of rRNA. J Clin Microbiol 28:1751–1759
Bouakaze C, Keyser C, de Martino SJ et al (2010) Identification and genotyping of Mycobacterium tuberculosis complex species using a SNaPshot minisequencing-based assay. J Clin Microbiol 48:1758–1766
Brosch R, Gordon SV, Marmiesse M et al (2002) A new evolutionary scenario for the Mycobacterium tuberculosis complex. Proc Natl Acad Sci USA 99:3684–3689
Brudey K, Gutierrez MC, Vincent V et al (2004) Mycobacterium africanum genotyping using novel spacer oligonucleotides in the direct repeat locus. J Clin Microbiol 42:5053–5057
Brudey K, Driscoll J, Rigouts L et al (2006) Mycobacterium tuberculosis complex genetic diversity: mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. BMC Microbiol 6:23
Carricajo A, Vincent V, Berthelot P et al (1999) Mycobacterial cross-contamination of bronchoscope detected by molecular techniques. J Hosp Infect 42:252–253
Cole ST, Brosch R, Parkhill J et al (1998) Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. Nature 393:537–544
Comas I, Gagneux S (2009) The past and future of tuberculosis research. PLoS Pathog 5(10):e1000600
Comas I, Homolka S, Niemann S et al (2009) Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. PLoS One 4:e7815
Coros A, DeConno E, Derbyshire KM (2008) IS6110, a Mycobacterium tuberculosis complex-specific insertion sequence, is also present in the genome of Mycobacterium smegmatis, suggestive of lateral gene transfer among mycobacterial species. J Bacteriol 190:3408–3410
Cousins DV, Skuce RA, Kazwala RR et al (1998) Towards a standardized approach to DNA fingerprinting of Mycobacterium bovis. International union against tuberculosis and lung disease, tuberculosis in animals subsection. Int J Tuberc Lung Dis 2:471–478
Daniel TM (2006) The history of tuberculosis. Respir Med 100:1862–1870
de Boer AS, Blommerde B, de Haas PEW et al (2002) False-positive mycobacterium tuberculosis cultures in 44 laboratories in The Netherlands (1993 to 2000): incidence, risk factors, and consequences. J Clin Microbiol 40:4004–4009
Del Portillo P, Murillo LA, Patarroyo ME (1991) Amplification of a species-specific DNA fragment of Mycobacterium tuberculosis and its possible use in diagnosis. J Clin Microbiol 29:2163–2168
Diaz R, Kremer K, de Haas PE et al (1998) Molecular epidemiology of tuberculosis in Cuba outside of Havana, July 1994-June 1995: utility of spoligotyping versus IS6110 restriction fragment length polymorphism. Int J Tuberc Lung Dis 2:743–750
Evans JT, Smith EG, Banerjee A et al (2007) Cluster of human tuberculosis caused by Mycobacterium bovis: evidence for person-to-person transmission in the UK. Lancet 369:1270–1276
Fabre M, Koeck JL, Le Fleche P et al (2004) High genetic diversity revealed by variable-number tandem repeat genotyping and analysis of hsp65 gene polymorphism in a large collection of" Mycobacterium canettii" strains indicates that the M. tuberculosis complex is a recently emerged clone of "M. canettii". J Clin Microbiol 42:3248–3255
Filliol I, Driscoll JR, Van Soolingen D et al (2002) Global distribution of Mycobacterium tuberculosis spoligotypes. Emerg Infect Dis 8:1347–1349
Filliol I, Driscoll JR, van Soolingen D et al (2003) Snapshot of moving and expanding clones of Mycobacterium tuberculosis and their global distribution assessed by spoligotyping in an international study. J Clin Microbiol 41:1963–1970

Filliol I, Motiwala AS, Cavatore M et al (2006) Global phylogeny of Mycobacterium tuberculosis based on single nucleotide polymorphism (SNP) analysis: insights into tuberculosis evolution, phylogenetic accuracy of other DNA fingerprinting systems, and recommendations for a minimal standard SNP set. J Bacteriol 188:759–772

Fleischmann RD, Alland D, Eisen JA et al (2002) Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains. J Bacteriol 184:5479–5490

Frieden TR, Sherman LF, Maw KL et al (1996) A multi-institutional outbreak of highly drug-resistant tuberculosis: epidemiology and clinical outcomes. JAMA 276:1229–1235

Friedman CR, Stoeckle MY, Johnson WD et al (1995) Double-repetitive-element PCR method for subtyping Mycobacterium tuberculosis clinical isolates. J Clin Microbiol 33:1383–1384

Frothingham R, Meeker-O'Connell WA et al (1998) Genetic diversity in the Mycobacterium tuberculosis complex based on variable numbers of tandem DNA repeats. Microbiology 144:1189–1196

Gagneux S, DeRiemer K, Van T et al (2006) Variable host–pathogen compatibility in Mycobacterium tuberculosis. Proc Natl Acad Sci USA 103:2869–2873

Ghanekar K, McBride A, Dellagostin O et al (1999) Stimulation of transposition of the Mycobacterium tuberculosis insertion sequence IS6110 by exposure to a microaerobic environment. Mol Microbiol 33:982–993

Gordon SV, Bottai D, Simeone R et al (2009) Pathogenicity in the tubercle bacillus: molecular and evolutionary determinants. Bioessays 31:378–388

Gutacker MM, Mathema B, Soini H et al (2006) Single-nucleotide polymorphism-based population genetic analysis of Mycobacterium tuberculosis strains from 4 geographic sites. J Infect Dis 193:121–128

Gutierrez MC, Brisse S, Brosch R et al (2005) Ancient origin and gene mosaicism of the progenitor of Mycobacterium tuberculosis. PLoS Pathog 1(1):e5

Haas WH, Butler WR, Woodley CL et al (1993) Mixed-linker polymerase chain reaction: a new method for rapid fingerprinting of isolates of the Mycobacterium tuberculosis complex. J Clin Microbiol 31:1293–1298

Hermans PW, van Soolingen D, van Embden JD (1992) Characterization of a major polymorphic tandem repeat in Mycobacterium tuberculosis and its potential use in the epidemiology of Mycobacterium kansasii and Mycobacterium gordonae. J Bacteriol 174:4157–4165

Hershberg R, Lipatov M, Small PM et al (2008) High functional diversity in Mycobacterium tuberculosis driven by genetic drift and human demography. PLoS Biol 6(12):e311

Hirsh AE, Tsolaki AG, DeRiemer K et al (2004) Stable association between strains of Mycobacterium tuberculosis and their human host populations. Proc Natl Acad Sci USA 101:4871–4876

Hocking TL, Choi C (1997) Tuberculosis: a strategy to detect and treat new and reactivated infections. Geriatrics 52:52–54

Huard RC, Fabre M, de Haas P et al (2006) Novel genetic polymorphisms that further delineate the phylogeny of the Mycobacterium tuberculosis complex. J Bacteriol 188:4271–4287

Joshi R, Reingold AL, Menzies D et al (2006) Tuberculosis among health-care workers in low- and middle-income countries: a systematic review. PLoS Med 3(12):e494

Kamerbeek J, Schouls L, Kolk A et al (1997) Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. J Clin Microbiol 35:907–914

Kaplan G, Post FA, Moreira AL et al (2003) Mycobacterium tuberculosis growth at the cavity surface: a microenvironment with failed immunity. Infect Immun 71:7099–7108

Kong Y, Cave MD, Zhang L et al (2006) Association between Mycobacterium tuberculosis Beijing/W lineage strain infection and extrathoracic tuberculosis: insights from epidemiologic and clinical characterization of the three principal genetic groups of M. tuberculosis clinical isolates. J Clin Microbiol 45:409–414

Kremer K, van Soolingen D, Frothingham R et al (1999) Comparison of methods based on different molecular epidemiological markers for typing of Mycobacterium tuberculosis complex strains: interlaboratory study of discriminatory power and reproducibility. J Clin Microbiol 37:2607–2618

Lai CC, Tan CK, Lin SH et al (2010) Clinical and genotypic characteristics of extensively drug-resistant and multidrug-resistant tuberculosis. Eur J Clin Microbiol Infect Dis 29:597–600

Le Flèche P, Fabre M, Denoeud F et al (2002) High resolution, on-line identification of strains from the Mycobacterium tuberculosis complex based on tandem repeat typing. BMC Microbiol 2:37

Li H, Ulstrup JC, Jonassen TO et al (1993) Evidence for absence of the MPB64 gene in some substrains of Mycobacterium bovis BCG. Infect Immun 61:1730–1734

López-Calleja AI, Gavín P, Lezcano MA et al (2009) Unsuspected and extensive transmission of a drug-susceptible Mycobacterium tuberculosis strain. BMC Pulm Med 9:3

Martín A, Herránz M, Serrano MJR et al (2007) Rapid clonal analysis of recurrent tuberculosis by direct MIRU-VNTR typing on stored isolates. BMC Microbiol 7:73

Martín A, Herranz M, Lirola MM et al (2008) Optimized molecular resolution of cross-contamination alerts in clinical mycobacteriology laboratories. BMC Microbiol 8:30

Martínez M, García de Viedma D, Alonso M et al (2006) Impact of laboratory cross-contamination on molecular epidemiology studies of tuberculosis. J Clin Microbiol 44:2967–2969

Mathema B, Kurepina NE, Bifani PJ et al (2006) Molecular epidemiology of tuberculosis: current insights. Clin Microbiol Rev 19:658–685

Mazars E, Lesjean S, Banuls AL et al (2001) High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. Proc Natl Acad Sci USA 98:1901–1906

Menzies D, Joshi R, Pai M (2007) Risk of tuberculosis infection and disease associated with work in health care settings. Int J Tuberc Lung Dis 11:593–605

Munsiff SS, Nivin B, Sacajiu G et al (2003) Persistence of a highly resistant strain of tuberculosis in New York city during 1990–1999. J Infect Dis 188:356–363

Niemann S, Köser CU, Gagneux S et al (2009) Genomic diversity among drug sensitive and multidrug resistant isolates of Mycobacterium tuberculosis with identical DNA fingerprints. PLoS One 4(10):e7407

Parwati I, Alisjahbana B, Apriani L et al (2010) Mycobacterium tuberculosis Beijing genotype is an independent risk factor for tuberculosis treatment failure in Indonesia. J Infect Dis 201:553–557

Pinsky BA, Banaei N (2008) Multiplex real-time PCR assay for rapid identification of Mycobacterium tuberculosis complex members to the species level. J Clin Microbiol 46:2241–2246

Post FA, Willcox PA, Mathema B et al (2004) Genetic polymorphism in Mycobacterium tuberculosis isolates from patients with chronic multidrug-resistant tuberculosis. J Infect Dis 190:99–106

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology 151:653–663

Prod'hom G, Guilhot C, Gutierrez MC et al (1997) Rapid discrimination of Mycobacterium tuberculosis complex strains by ligation-mediated PCR fingerprint analysis. J Clin Microbiol 35:3331–3334

Roring S, Scott A, Brittain D et al (2002) Development of variable-number tandem repeat typing of Mycobacterium bovis: comparison of results with those obtained by using existing exact tandem repeats and spoligotyping. J Clin Microbiol 40:2126–2133

Ross BC, Raios K, Jackson K et al (1992) Molecular cloning of a highly repeated DNA element from Mycobacterium tuberculosis and its use as an epidemiological tool. J Clin Microbiol 30:942–946

Ruddy M, McHugh TD, Dale JW et al (2002) Estimation of the rate of unrecognized cross-contamination with Mycobacterium tuberculosis in London microbiology laboratories. J Clin Microbiol 40:4100–4104

Schürch AC, Kremer K, Kiers A et al (2010) The tempo and mode of molecular evolution of Mycobacterium tuberculosis at patient-to-patient scale. Infect Genet Evol 10:108–114

Schwoebel V, Decludt B, de Benoist AC et al (1998) Multidrug resistant tuberculosis in France 1992–1994: two case-control studies. BMJ 317:630–631

Small PM, McClenny NB, Singh SP et al (1993) Molecular strain typing of Mycobacterium tuberculosis to confirm cross-contamination in the mycobacteriology laboratory and modification of procedures to minimize occurrence of false-positive cultures. J Clin Microbiol 31:1677–1682

Small PM, Hopewell PC, Singh SP et al (1994) The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. N Engl J Med 330:1703–1709

Smittipat N, Billamas P, Palittapongarnpim M et al (2005) Polymorphism of variable-number tandem repeats at multiple loci in Mycobacterium tuberculosis. J Clin Microbiol 43:5034–5043

Sreevatsan S, Pan X, Stockbauer KE et al (1997) Restricted structural gene polymorphism in the Mycobacterium tuberculosis complex indicates evolutionarily recent global dissemination. Proc Natl Acad Sci USA 94:9869–9874

Sun Y, Bellamy R, Lee ASG et al (2004) Use of mycobacterial interspersed repetitive unit-variable-number tandem repeat typing to examine genetic diversity of Mycobacterium tuberculosis in Singapore. J Clin Microbiol 42:1986–1993

Sunder S, Lanotte P, Godreuil S et al (2009) Human-to-human transmission of tuberculosis caused by Mycobacterium bovis in immunocompetent patients. J Clin Microbiol 47:1249–1251

Supply P, Magdalena J, Himpens S et al (1997) Identification of novel intergenic repetitive units in a mycobacterial two-component system operon. Mol Microbiol 26:991–1003

Supply P, Mazars E, Lesjean S et al (2000) Variable human minisatellite-like regions in the Mycobacterium tuberculosis genome. Mol Microbiol 36:762–771

Supply P, Lesjean S, Savine E et al (2001) Automated high-throughput genotyping for study of global epidemiology of Mycobacterium tuberculosis based on mycobacterial interspersed repetitive units. J Clin Microbiol 39:3563–3571

Supply P, Allix C, Lesjean S et al (2006) Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. J Clin Microbiol 44:4498–4510

Talarico S, Donald Cave M, Foxman B et al (2007) Association of Mycobacterium tuberculosis PE_PGRS33 polymorphism with clinical and epidemiological characteristics. Tuberculosis 87:338–346

Talarico S, Zhang L, Marrs CF et al (2008) Mycobacterium tuberculosis PE_PGRS16 and PE_PGRS26 genetic polymorphism among clinical isolates. Tuberculosis 88:283–294

Thierry D, Brisson-Noël A, Vincent-Lévy-Frébault V et al (1990a) Characterization of a Mycobacterium tuberculosis insertion sequence, IS6110, and its application in diagnosis. J Clin Microbiol 28:2668–2673

Thierry D, Cave MD, Eisenach KD et al (1990b) IS6110, an IS-like element of Mycobacterium tuberculosis complex. Nucleic Acids Res 18:188

Thwaites G, Caws M, Chau TTH et al (2008) Relationship between Mycobacterium tuberculosis genotype and the clinical phenotype of pulmonary and meningeal tuberculosis. J Clin Microbiol 46:1363–1368

Tsolaki AG, Hirsh AE, DeRiemer K et al (2004) Functional and evolutionary genomics of Mycobacterium tuberculosis: insights from genomic deletions in 100 strains. Proc Natl Acad Sci USA 101:4865–4870

Tsolaki AG, Gagneux S, Pym AS, de la Salmoniere G et al (2005) Genomic deletions classify the Beijing/W strains as a distinct genetic lineage of Mycobacterium tuberculosis. J Clin Microbiol 43:3185–3191

Valway SE, Greifinger RB, Papania M et al (1994) Multidrug-resistant tuberculosis in the New York State prison system, 1990–1991. J Infect Dis 170:151–156

van Embden JD, Cave MD, Crawford JT et al (1993) Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 31:406–409

van Soolingen D, Hermans PW, de Haas PE et al (1992) Insertion element IS1081-associated restriction fragment length polymorphisms in Mycobacterium tuberculosis complex species: a reliable tool for recognizing Mycobacterium bovis BCG. J Clin Microbiol 30:1772–1777

van Soolingen D, Qian L, de Haas PE et al (1995) Predominance of a single genotype of Mycobacterium tuberculosis in countries of east Asia. J Clin Microbiol 33:3234–3238

Warren RM, Victor TC, Streicher EM et al (2004) Patients with active tuberculosis often have different strains in the same sputum specimen. Am J Respir Crit Care Med 169:610–614

WHO report (2009) Global tuberculosis control: a short update to the 2009 report. http://www.who.int/tb/publications/global_report/2009/update/en/index.html

Wiid IJ, Werely C, Beyers N et al (1994) Oligonucleotide (GTG)5 as a marker for Mycobacterium tuberculosis strain identification. J Clin Microbiol 32:1318–1321

Zhang Y, Heym B, Allen B et al (1992) The catalase-peroxidase gene and isoniazid resistance of Mycobacterium tuberculosis. Nature 358:591–593

Zhang J, Abadia E, Refregier G et al (2010) Mycobacterium tuberculosis complex CRISPR genotyping: improving efficiency, throughput and discriminative power of 'spoligotyping' with new spacers and a microbead-based hybridization assay. J Med Microbiol 59:285–294

Chapter 8
Pneumocystis Molecular Phylogeny: A Way to Understand Both Pneumocystosis Natural History and Host Taxonomy

Magali Chabé, Jean-Pierre Hugot, and Eduardo Dei-Cas

Abstract The genus *Pneumocystis* comprises numerous fungal pathogens dwelling in the lungs of a wide spectrum of mammalian species. Here, we show how new molecular methodologies, and in particular phylogenetics, have changed our views on the diversity and epidemiology of these fungal pathogens. *Pneumocystis* organisms have a marked host-species-related diversity, which is associated to close host specificity. High divergence among *Pneumocystis* species probably resulted from a prolonged process of coevolution with each mammal host and mostly associated with cospeciation like it was demonstrated in recent phylogenetic studies on primate-derived *Pneumocystis*. Therefore, *Pneumocystis* species or strains may be used as evolution markers, to reconstruct the phylogenetic history of mammal species, thus helping to elucidate controversial issues in mammalian taxonomy.

M. Chabé (✉)
Biology and Diversity of Emergent Eukaryotic Pathogens (BDEEP) – Center for Infection and Immunity of Lille, Pasteur Institute of Lille, Inserm U1019, CNRS UMR 8204, University Lille-Nord-de-France, Lille, France

Department of Parasitology – Mycology, Faculty of Biological and Pharmaceutical Sciences, University of Lille-Nord-de-France, Lille, France
e-mail: magali.chabe@univ-lille2.fr

J.-P. Hugot
OSEB, UMR 5202 CNRS, Muséum National d'Histoire Naturelle, Paris, France

E. Dei-Cas
Biology and Diversity of Emergent Eukaryotic Pathogens (BDEEP) – Center for Infection and Immunity of Lille, Pasteur Institute of Lille, Inserm U1019, CNRS UMR 8204, University Lille-Nord-de-France, Lille, France

Department of Parasitology – Mycology, Faculty of Medicine, University of Lille-Nord-de-France, Biology-Pathology Centre, University Hospital Center, Lille, France

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_8,

8.1 Introduction: New Views on the *Pneumocystis* Natural History

For the last years, important research advances changed radically our conceptions about *Pneumocystis* development, infection sources, reservoir, and taxonomy. Until recently it was considered that *Pneumocystis* organisms were able for multiplying exclusively in deeply immunodepressed hosts, that infection could be contracted by airborne route from animal or hypothetical environmental sources, where the unique species of the genus ("*Pneumocystis carinii*"), able to cause *Pneumocystis* pneumonia (PcP) in both humans and animals, developed saprophytic growth.

Recent researches have changed drastically this view. Sensitive molecular methods allowed the detection of *Pneumocystis* organisms in respiratory samples from healthy or hospitalized subjects without severe immunodepression (Calderon et al. 1996, 2004; Dei-Cas 2000; Peterson and Cushion 2005). Moreover, it was shown that *Pneumocystis* organisms were able to multiply transiently in the lungs of immunocompetent hosts, these hosts being able to both transmit the infection to susceptible hosts by airborne route and play the role of *Pneumocystis* reservoir (Chabé et al. 2004).

Furthermore, molecular genetic studies revealed a marked host-species related *Pneumocystis* genetic heterogeneity. Consistently, close host species specificity was demonstrated by cross infection experiments (Aliouat et al. 1993b, 1994; Gigliotti et al. 1993; Durand-Joly et al. 2002). On the basis of genomic and phenotypic divergence multiple *Pneumocystis* species were then recognized (Frenkel 1999; Dei-Cas 2000; Stringer et al. 2001; Cushion et al. 2004; Keely et al. 2004). *Pneumocystis jirovecii* Frenkel was the sole species identified in humans, showing also close host specificity (Durand-Joly et al. 2002). Gene sequence analysis and isoenzymatic data suggested that speciation in the group resulted from long genetic isolation and coevolution (Demanche et al. 2001; Hugot et al. 2003).

In the present chapter, we attempted to characterize, mainly from basic molecular phylogenetic studies developed for the last 20 years, the *Pneumocystis* genus, life cycle and transmission of *Pneumocystis* organisms, the taxonomy of the group at the species level, the close cophylogeny with mammals hosts, and the impact of new data on both host taxonomy and basic notions on *Pneumocystis* infection epidemiology.

8.2 A Short Presentation of the *Pneumocystis* Genus

On the basis of genomic and other data (see Aliouat-Denis et al. 2008 for review), the *Pneumocystis* genus was assigned to the group of fungi at the branch point between Ascomycota and Basidiomycota (Edman et al. 1988, 1989; Wakefield et al. 1992; Eriksson 1994; Guarro et al. 1999; Wakefield 2002; Frealle et al. 2005, 2006). Specifically, the genus was included in the taxon Taphrinomycotina (Eriksson and

Winka 1997), formerly "Archiascomycetes" (Nishida and Sugiyama 1993), which includes fission yeasts (*Schizosaccharomyces* spp), *Pneumocystis* and other genera (Liu et al. 2009). Indeed, Taphrinomycotina would diverge as a sister group to Saccharomycotina plus Pezizomycotina, as suggested by the recent work of Liu et al. (2009).

8.2.1 Host Range of *Pneumocystis* Genus

Pneumocystis (Delanoë and Delanoë 1912) constitutes a highly diversified eukaryotic genus spanning numerous parasitic species that are host species specific and well adapted to live inside the lungs of a great diversity of terrestrial, aerial and, likely, aquatic mammals (Settnes and Henriksen 1989; Laakkonen et al. 1993, 2001; Peters et al. 1994b; Lobetti et al. 1996; Bishop et al. 1997; Laakkonen and Sukura 1997; Mazars et al. 1997a; Wakefield et al. 1997; Laakkonen 1998; Perron-Lepage et al. 1999; Guillot et al. 1999, 2001; Durand-Joly et al. 2000; Demanche et al. 2001, 2003; English et al. 2001; Hugot et al. 2003; Dei-Cas et al. 2006; Sanches et al. 2007; Chabé et al. 2010).

Hitherto, microscopic or molecular methods allowed the detecting of *Pneumocystis* in all explored mammals with the exception of Monotremes (Table 8.1). Two studies searched for *Pneumocystis* in birds with apparent contradictory results. Thus, a survey of animals in Japan reported the presence of *Pneumocystis* in chicken (5%), quails (17%) and feral pigeons (29%) by using microscopy and a cyst concentration method (Shimizu et al. 1985). In contrast, Poelma (1975) was unable to detect *Pneumocystis* organisms in Zoo birds using light microscopy. Likewise, Settnes et al. (1994) did not detect these fungi in European birds by using either microscopic lung smear examination or PCR on serum samples, although they were able to detect *Pneumocystis* serum antibody in some specimens.

8.2.2 Morphology of *Pneumocystis*

Pneumocystis organisms are extracellular microbes able to attach intimately to alveolar epithelial cells (Dei-Cas et al. 2004), and occasionally, to induce severe pneumonitis, particularly in hosts with marked impairment of the immune system (Dei-Cas 2000). In such hosts, *Pneumocystis* cells proliferate progressively and may fill pulmonary alveolar cavities, a process leading to respiratory failure.

Ultrastructural studies have shown that *Pneumocystis* organisms, especially the mononucleate trophic forms (formerly called 'trophozoites') (Fig. 8.1), attach specifically to type-1 epithelial alveolar cells. Trophic forms (2–8 μm in diameter) have a thin cell wall and emit cytoplasmic expansions or filopodia that vary in thickness depending on the species (Dei-Cas et al. 1994, 2004; Mazars and Dei-Cas 1998; Nielsen et al. 1998). Filopodia may penetrate deeply into the cytoplasm of the host

Table 8.1 *Pneumocystis* occurrence in mammals

Mammalian hosts		Geographical origin	*Pneumocystis* species[a]	Host origin (laboratory, zoo, domestic or wild animals)	PcP or mild infection	*Pneumocystis* targetted genes	Authors
Rodentia	Rat (*Rattus norvegicus*)	Old and New World	*P. carinii*	Laboratory and wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, ITS, TS, β-tub, α-tub, ATPase , TBP, MSG, HSP70, DHPS, DHFR, arom, MnSOD, cdc2, cyt b	See Aliouat-Denis et al. (2008) for review
	Rat (*Rattus norvegicus*)	Old and New World	*P. wakefieldiae*	Laboratory and wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, ITS, TS, ATPase, TP, MSG, HSP70	See Aliouat-Denis et al. (2008) for review
	Rat (*Rattus rattus*)	Denmark, Japan	–	Bred in captivity and wild animals	ND	None	Settnes and Lodal (1980); Laakkonen (1998)
	Dusky-footed rat (*Neotoma fuscipes*)	California	–	Wild animals	Mild infection	None	Laakkonen et al. (2001)
	Mouse (*Mus musculus*)	Old and New World	*P. murina*	Laboratory and wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, ITS, TS, HSP70, DHPS, DHFR, MSG, arom, MnSOD, cdc2	See Aliouat-Denis et al. (2008) for review

Guinea pig (*Cavia porcellus*)	Old and New World	–	Laboratory animals	PcP, mild infection	None	Chagas (1909); Yoshida et al. (1981); Reed and O'Donoghue (1979)
Agouti paca	French Guyana	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
Proechimys guyanensis	French Guyana	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
Apodemus spp	Europe, Japan	–	Wild animals	Mild infection	mtLSUrRNA (in *Apodemus sylvaticus*)	See Laakkonen (1998) for review, Mazars et al. (1997a)
Micromys minutes	Finland	–	Wild animals	ND	None	Laakkonen (1998)
California meadow vole (*Chaetodipus californicus)*	California	–	Wild animals	Mild infection	None	Laakkonen et al. (2001)
Voles (*Clethrionomys* spp)	Czech Republic, Finland	–	Wild animals maintained or not in captivity	Mild infection	None	See Laakkonen (1998) for review
Voles (*Microtus* spp)	Europe, Japan, California	–	Wild animals maintained or not in captivity	Mild infection	mtLSUrRNA (in *Microtus agrestis*)	See Laakkonen (1998) for review, Laakkonen et al. (2001); Mazars et al. 1997a

(continued)

Table 8.1 (continued)

Mammalian hosts		Geographical origin	*Pneumocystis* species[a]	Host origin (laboratory, zoo, domestic or wild animals)	PcP or mild infection	*Pneumocystis* targetted genes	Authors
	European water vole (*Arvicola terrestris*)	France	–	Wild animals	Mild infection	None	Goyot et al. (1986)
	Oryzomys capito	Brazil	–	Wild animals	ND	None	Lainson and Shaw (1975)
	Garden dormouse (*Eliomys quercinus*)	France	–	Wild animals	Mild infection	mtLSUrRNA	Mazars et al. (1997a)
	Gerbil (*Meriones unguiculatus)*	China	–	Laboratory animals	ND	NucrRNA, ITS	Li et al. (2008)
	Squirrel (*Sciurus aestuans*)	French Guyana	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
Lagomorpha	Rabbit (*Oryctolagus cuniculus*)	Old World	*P. oryctolagi*	Meat, laboratory and wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, ITS, TS, HSP70, DHPS, DHFR, arom, mnSOD	See Dei-Cas et al. (2006) for review
	Hare (*Lepus europaeus*)	Denmark, Finland	–	Wild animals	PcP, mild infection	None	Settnes et al. (1986); Laakkonen et al. (2006); Blazec (1960)
	Hare (*Lepus timidus*)	Finland	–	Wild animals	PcP, mild infection	None	Laakkonen et al. (2006)

Soricomorpha	Shrews (*Sorex* spp)	Old and New World	–	Wild animals	Mild infection	mtLSUrRNA (in *Sorex araneus*)	Laakkonen (1998); Laakkonen et al. (2001); Mazars et al. (1997a); Peters et al. (1994a)
	Desert shrew (Notiosorex crawfordi)	California	–	Wild animals	Mild infection	None	Laakkonen et al. (2001)
	Suncus murinus	Taiwan	–	Wild animals maintained in captivity	Mild infection	None	Cross and Hung (1975)
	Mole (*Talpa europaea*)	Spain, France	–	Wild animals	Mild infection	mtLSUrRNA	Mazars et al. (1997a)
Afrosoricida	Tenrec (*Migrogale* sp)	Madagascar	–	Wild animals	Mild infection	None	Laakkonen (1998)
Carnivora	Ferret (*Mustela* spp)	Old and New World	–	Laboratory and wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, TS, MSG, arom, DHFR, DHPS	Laakkonen (1998); See Aliouat-Denis et al. (2008) for review
	Fox (*Vulpes vulpes*)	Denmark	–	Wild animals	ND	None	Settnes et al. (1986)
	Fennec fox (*Vulpes zerda*)	Netherlands	–	Zoo animals	ND	None	Poelma (1975)
	Dog (*Canis lupus vulgaris*)	Old and New World	–	Domestic animals	PcP	mtLSUrRNA, mtSSUrRNA, DHFR, DHPS, TS	Guillot et al. (2001); Sukura et al. (1996); Aliouat-Denis et al. (2008)

(continued)

Table 8.1 (continued)

Mammalian hosts		Geographical origin	*Pneumocystis* species[a]	Host origin (laboratory, zoo, domestic or wild animals)	PcP or mild infection	*Pneumocystis* targetted genes	Authors
	Cat (*Felis catus*)	Denmark, Japan	–	Domestic animals	Mild infection	None	Settnes and Hasselager (1984); Shiota et al. (1990)
	Grison (*Galictis vittata*)	French Guyana	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
	Coatimundi (*Nasua narica*)	Brazil	–	Wild animals maintained in captivity	PcP	None	Lainson and Shaw (1975)
	Red panda (*Ailurus fulgens*)	Netherlands	–	Zoo animals	ND	None	Poelma (1975)
Xenarthra	Sloth (*Bradypus tridactylus*)	Brazil	–	Wild animals maintained in captivity	PcP	None	Lainson and Shaw (1975)
	Sloth (*Choloepus didactylus*)	Brazil	–	Wild animals	ND	None	Lainson and Shaw (1975)
	Sloth (*Bradypus variegatus*)	Panama	–	Wild animal maintained in captivity	PcP	None	Yonushonis et al. (1986)
Perissodactyla	Horse (*Equus cabalus*)	England, France	–	Domestic animals	PcP	mtLSUrRNA	Peters et al. (1994b); Perron-Lepage et al. (1999)
Artiodactyla	Roe deer (*Capreolus capreolus*)	Denmark	–	Wild animals	ND	None	Settnes et al. (1986)

	Goat (*Capra aegagrus*)	South Africa	–	Domestic animal	PcP	None	Mc Connell et al. (1971)
	Sheep (*Ovis aries*)	Denmark	–	Domestic animals	ND	None	Settnes and Henriksen (1989)
	Calf (*Bovis taurus*)	Denmark	–	Domestic animals	ND	None	Settnes and Henriksen (1989)
	Sable antelope (*Hippotragus niger*)	South Africa	–	Wild animals	Mild infection	None	Wilson et al. (1974)
	Boar (*Sus crofa*)	Europe	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
	Pig (*Sus domesticus*)	Old and New World	–	Domestic animals	PcP, mild infection	mtLSUrRNA, MnSOD	Sanches et al. (2007); Kondo et al. (1993); Settnes et al. (1991); Aliouat-Denis et al. (2008)
Didelphimorphia (Marsupials)	*Marmosa murina*	French Guyana	–	Wild animals	ND	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001)
Diprotodontia (Marsupials)	Red kangaroo (*Macropus rufus*)	Netherlands	–	Zoo animals	ND	None	Poelma (1975)
Chiroptera	*Glossophaga sorocina*	French Guyana, Mexico	–	Wild animals	Mild infection	mtLSUrRNA, mtSSUrRNA	Guillot et al. (2001); Derouiche et al. (2009)

(continued)

Table 8.1 (continued)

Mammalian hosts		Geographical origin	*Pneumocystis* species[a]	Host origin (laboratory, zoo, domestic or wild animals)	PcP or mild infection	*Pneumocystis* targetted genes	Authors
	Pipistrellus pipistrellus	France	–	Wild animals	Mild infection	mtLSUrRNA, mtSSUrRNA	Derouiche et al. (2009)
	Tadarida brasiliensis	Mexico, Argentina, Brazil	–	Wild animals	Mild infection	mtLSUrRNA, mtSSUrRNA	Derouiche et al. (2009); Cavallini-Sanches et al. (2009)
	Nyctinomops laticaudatus, *Desmodus rotundus*, *Molossus molossus*, *Artibeus fimbriatus*, *Sturnira lilium*, *Myotis levis*, *Diphylla ecaudata*	Brazil	–	Wild animals	ND	mtLUSrRNA, mtSSUrRNA	Cavallini-Sanches et al. (2009)
Primates	Man (*Homo sapiens*)	Old and New World	*P. jirovecii*		PcP, mild infection	mtLSUrRNA, mtSSUrRNA, nucrRNA, ITS, TS, β-tub, α-tub, HSP70, DHPS, DHFR, arom, MSG, MnSOD, cdc2, cyt b	See Aliouat-Denis et al. (2008) for review

Non-human Primates	Old and New World	–	Zoo, laboratory or wild animals	PcP, mild infection	mtLSUrRNA, mtSSUrRNA, DHPS, nucrRNA, MSG, ITS, MnSOD (in *Macaca mulatta*),DHFR (in *Aotus nancymai*), None[b]	See Aliouat-Denis et al. (2008) and Durand-Joly et al. (2000)

mtLSUrRNA large subunit of mitochondrial ribosomal RNA; *mtSSUrRNA* small subunit of mitochondrial ribosomal RNA, nucrRNA, ribosomal RNA from nuclear genome; *ITS* internal transcribed spacer from nuclear rRNA gene locus; *TS* thymidylate synthase; *ATPase* proton pump; *TBP* TATA-box binding protein; *MSG* major surface glycoprotein genes; *arom* gene encoding a single polypeptide that catalyzes five consecutive steps in prechorismate aromatic amino acid biosynthesis; *DHFR* dihydrofolate reductase; *DHPS* dihydropteorate synthase; *HSP70* heat shock protein 70; *MnSOD* manganese superoxide dismutase; *cdc2* cell division cycle 2; *cytb* cytochrome b

[a]Only formally described *Pneumocystis* species were indicated

[b]*Pneumocystis* m*tLSUrRNA* gene was amplified in: *Cercopithecus lhoesti, Macaca nemestrina, Macaca fascicularis, Microcebus murinus, Saguinus oedipus oedipus, Callithrix jacchus, Saguinus midas midas, Callithrix geoffroyi, Cercopithecus hamlyni, Saimiri sciureus, Hapalemur griseus, Cercopithecus nictitans, Callimico goeldii, Saguinus fuscicollis, Maki macaco, Allenopithecus nigroviridis, Saguinus imperator,* Chinese macaque rhesus, Indian macaque rhesus, *Pithecia pithecia*

Pneumocystis mtSSUrRNA gene was amplified in: Indian macaque rhesus, *Macaca fascicularis, Cercopithecus hamlyni, Cercopithecus nictitans, Callithrix jacchus, Callithrix geoffroyi, Saguinus midas midas, Saguinus fuscicolis, Saguinus oedipus oedipus, Callimico goeldii, Pithecia pithecia, Saimiri sciureus, Maki macaco*

Pneumocystis DHPS gene was amplified in: *Allenopithecus nigroviridis, Callithrix geoffroyi, Saimiri sciureus, Callimico goeldii, Saguinus fuscicollis, Saguinus midas midas*

Pneumocystis organisms were detected in: *Macaca fuscata fuscata, Macaca cyclopis, Aotus trivirgatus, Pan troglodytes, Galago senegalensis, Galago demidovii, Alouatta fusca, Lagothrix lagotricha, Ateles belzebuth, Callithrix aurita, Callithrix penicillata*

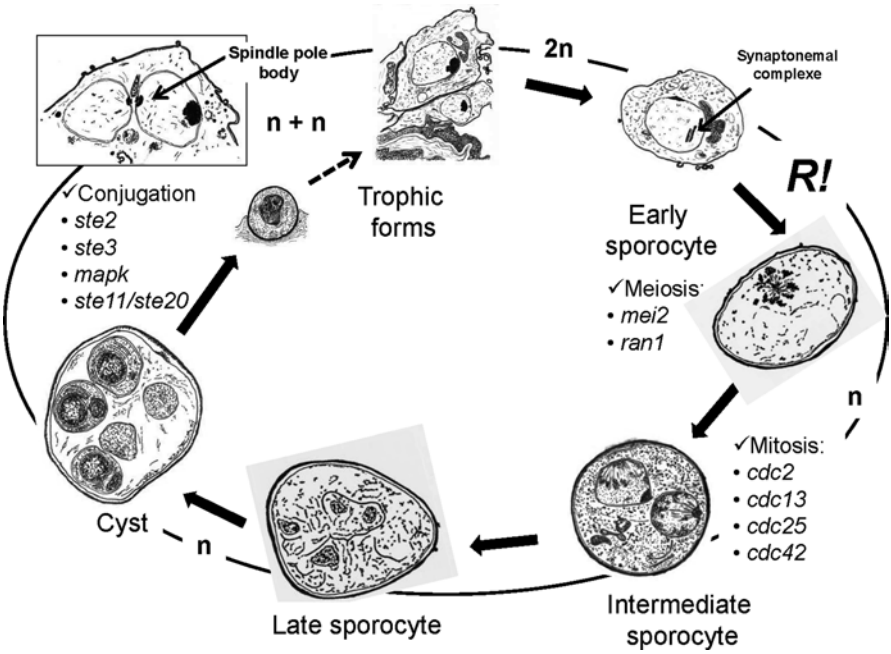

Fig. 8.1 Hypothetical life-cycle of *Pneumocystis* species (from Chabé et al. 2011 in FEMS Yeast Res). Parasites are represented as observed in the lung using transmission electron microscopy. Pleomorphic, thin-walled mononuclear trophic forms are shown attached to type-I epithelial alveolar cells (*at the top*). Trophic forms (*2n*) evolve into early sporocyte in which a synaptonemal complex is indicated. Meiotic nuclear division (*R!*) leads to thickwalled sporocytic and cystic stages, in which multiple nuclear divisions in intermediate and late sporocytes lead to the formation of eight haploid spores or ascospores (*n*). These forms are able to leave the cyst, to attach specifically to type-I epithelial alveolar cells and, likely, to develop conjugation as illustrated at the top (*left*) (*n* + *n*), where spindle pole bodies are well visible. Main molecular factors involved in conjugation, meiosis or mitosis identified in *Pneumocystis carinii* have been listed in the figure (see main references in the text). Synaptonemal complex and conjugation pictures have been drawn according to ultrastructural micrographies of Yoshida et al. (1984) and Itatani (1996), respectively

cell (Dei-Cas et al. 1991). However, no disruption of host cell membrane results from either attachment or filopodial activity. In addition, no structural or functional host cell alteration was found in in vivo or in vitro studies using transmission electron microscopy (TEM) (Dei-Cas et al. 1991, 2004; Settnes and Nielsen 1991; Aliouat et al. 1993a), confocal microscopy (unpublished) or exploring the alveolar epithelium cytophysiology (Beck et al. 1998). Trophic forms coexist usually with cystic forms (4–8 μm in diameter) that are spheroid structures with a thick cell wall, likely involved in airborne transmission (Creusy et al. 1996; Dei-Cas 2000; Cushion et al. 2010).

8.2.3 *Atypical Features of Pneumocystis*

In clinical or research laboratory, *Pneumocystis* organisms are often considered as atypical fungi (Cailliez et al. 1996). Actually, though of fungal nature, they are unable to grow in vitro in current fungal culture media, they have not ergosterol in their plasma membrane, and they are not susceptible to most antifungal drugs, excepting echinocandins, which target beta-glucan synthase (Powles et al. 1998; Ito et al. 2000; Utili et al. 2007). The absence of ergosterol can explain why *Pneumocystis* organisms are neither susceptible to inhibitors of ergosterol synthesis nor to amphotericin B, though this drug apparently binds to *Pneumocystis* sterols in vitro and causes permeability changes in the parasite cellular membrane (Kaneshiro et al. 2000).

Like other microfungi, *Pneumocystis* organisms possess chitin and beta-glucan in their cell wall. For this reason, the cystic forms, which have a thick cell wall, can be stained with calcofluor, ortho-toluidine blue, PAS or Gomori-Grocott's or similar silver stains (Dei-Cas et al. 2004, 2006). However, all known *Pneumocystis* life cycle stages present also an outer membrane that resembles that of Gram-negative prokaryotes and cyanobacteria (Barton and Campbell 1967; Vavra and Kucera 1970; Campbell 1972; Haque et al. 1987; De Stefano et al. 1990; Palluault et al. 1992; Dei-Cas et al. 2004). This intriguing structure is apparently absent from the cell wall of other fungi.

Another striking atypical *Pneumocystis* feature is the presence of only one rRNA gene (Giuntoli et al. 1994; Tang et al. 1998) while the vast majority of eukaryotes, including fungi, possess dozens of copies, often hundreds of rRNA genes (Fischer et al. 2006). Interestingly, despite its low copy number, *Pneumocystis* rRNA gene evolved at a rate typical for eukaryotes, as emerged from a 18S rDNA comparative analysis (Fischer et al. 2006).

8.3 *Pneumocystis* Life-Cycle and Transmission

Lack of robust, long-term culture has prevented detailed dynamic follow-up of differentiation of the *Pneumocystis* life cycle stages. Therefore, hypotheses on the *Pneumocystis* life-cycle stem mainly from ultrastructural and molecular data as well as from in vivo observations on the growth rate (Aliouat et al. 1999; Aliouat-Denis et al. 2009). Such hypotheses were recently reviewed in detail by Aliouat-Denis et al. (2009). In short, life cycle views include both sexual and asexual reproduction alternate (Yoshida et al. 1984; Yoshida 1989; Cushion 2004; Dei-Cas et al. 2004; De Souza and Benchimol 2005; Thomas and Limper 2007; Aliouat-Denis et al. 2008, 2009). *Pneumocystis* life cycle stages are trophic forms, sporocytes and mature cysts (Fig. 8.1). Trophic forms are the most abundant stages, representing 90–95% of the parasite population in the lungs of hosts with *Pneumocystis* pneumonia (PcP). They are mononuclear, mostly haploid cells (Cornillot et al. 2002; Dei-Cas et al. 2004) presenting a single-layered, electron-dense thin cell wall. Ameboid in shape, trophic forms display filopodia. Ultrastructural observations (Itatani 1996) as well

as the recent discovery of the *Ste3* pheromone receptor gene and the *Ste2*-like homologue within the *Pneumocystis* genome suggest the existence of mating types and the occurrence of conjugation between trophic forms (Smulian et al. 2001; Cushion 2004) leading to the early sporocyte, a round, thin-walled, mononuclear diploid cell. Shortly thereafter early sporocyte enters a meiotic division process as attested by the detection of synaptonemal complexes (Matsumoto and Yoshida 1984; Peters et al. 2001). Additional mitotic replications result in eight nuclei in the sporocyte stage (Fig. 8.1) in whose cell wall develops an internal, beta-glucan rich electron-lucent layer. Late sporocyte evolves then into thick-walled mature cyst (Dei-Cas et al. 2004) with few filopodia. In the mature cyst, are generated eight haploid ascospores, containing a single nucleus and a fairly dense cytoplasm with a rounded mitochondrion and a well-developed rough endoplasmic reticulum (Dei-Cas et al. 2004). Ascospores leave then the cyst or ascus, likely through a foramen-like structure (Itatani 1994), to develop into eight free haploid trophic forms able to attach to the alveolar epithelium and restart the cycle (Yoshida 1989). Many factors involved in both conjugation and meiosis have been identified in *Pneumocystis carinii* for the last years (Fig. 8.1) (Thomas et al. 1998; Wyder et al. 1998; Kottom et al. 2000, 2003; Smulian et al. 2001; Vohra et al. 2003a, b, 2004; Cushion 2004; Cushion et al. 2007; Burgess et al. 2008; Villegas et al. 2009).

Interestingly, *Pneumocystis* species develop sexual reproduction within their host (Dei-Cas et al. 1992; Cushion et al. 2007) as well as biotrophic fungi of plants (O'Connell and Panstruga 2006). In contrast, most fungi developing in animals do not appear to complete a sexual cycle (Sexton and Howlett 2006). Some exceptions to this rule may nevertheless be noted. *Aspergillus nidulans* is able to sexually reproduce in vivo in human as well as in animal tissues, thus producing cleistothecia and Hülle cells (Doby and Kombila-Favry 1978; Dei-Cas and Vernes 1986; Mitchell et al. 1987). More recently, it was shown that clinical *Candida lusitaniae* isolates were able to mate and undergo meiosis once put in contact with a sexually compatible strain (François et al. 2001).

Likely, the most striking singularity of *Pneumocystis* organisms as fungal pathogens to animals (Dei-Cas 2000) is inter-host airborne transmission. Actually, while *Pneumocystis* organisms can be directly transmitted among hosts by the airborne route (see Aliouat-Denis et al. 2008, for review and original references), it is well known that deep fungal diseases caused by typical dimorphic, exosaprophytic fungi such as *Histoplasma capsulatum* or *Coccidioides immitis*, cannot be transmitted from one infected host to another. The reservoir for these pathogens is the soil, which is also the infection source for hosts (Dei-Cas and Vernes 1986). Thus, PcP would be the sole deep fungal disease transmissible directly by airborne route between hosts (Dei-Cas 2000). Furthermore, whereas the known *Pneumocystis* species or strains show close host species specificity (Aliouat et al. 1993b, 1994; Gigliotti et al. 1993; Durand-Joly et al. 2002), no evidence of narrow host specificity was reported in other agents of deep mycosis in animals (Dei-Cas 2000; De Hoog et al. 2000; Sexton and Howlett 2006).

Furthermore, *Pneumocystis* in utero transmission seems to occur in some mammal hosts like lagomorphs and primates, including humans (Ceré et al. 1997;

Sanchez et al. 2007; Montes-Cano et al. 2009). This modality of transmission does not seem to be usual in other pathogenic fungi, though fungal infection is a major worldwide cause of abortion in cattle (Kirkbride 1992).

8.4 The *Pneumocystis* Species

As evoked above, molecular tools have been crucial for the assigning of *Pneumocystis* genus to the group of Fungi. Molecular approaches, specifically genetic studies aiming at exploring potential divergence between *Pneumocystis* strains isolated from different mammal species, have also been critical to identify such strains, and to assess both divergence among them and stability in time and space (Dei-Cas et al. 1998). Indeed, as the *Pneumocystis* isolates from different mammals were morphologically indistinguishable (at least at the level of light microscopy), it was considered for almost one century that the *Pneumocystis* genus contained a unique widely spread euryxenic species. However, in the 90', a cluster of host species-related genomic (Sinclair et al. 1991; Wakefield et al. 1992; Stringer et al. 1993; Banerji et al. 1995; Mazars et al. 1995; Ortiz-Rivera et al. 1995; Hunter and Wakefield 1996; Stedman et al. 1998; Denis et al. 2000; Demanche et al. 2001; Ma et al. 2001), karyotypic (Weinberg and Bartlett 1991; Stringer et al. 1993), isoenzymatic (Dei-Cas et al. 1994; Mazars et al. 1994, 1997b) and antigenic differences (Furuta and Ueda 1987; Linke et al. 1989; Gigliotti 1992; Bauer et al. 1993; Christensen et al. 1996) were reported among *Pneumocystis* isolates from diverse mammals. Thus, the need of considering them like substantially different taxons emerged clearly, and the species level was considered to class these new entities. Furthermore, host species-related genomic divergence in the *Pneumocystis* genus was often higher than that existing among typical ascomycetous species (Dei-Cas et al. 1998; Stringer et al. 2001; Keely et al. 2003).

The Phylogenetic Species Concept (PSC) as specified by Taylor et al. (2000) for fungi, revealed operative to describe *Pneumocystis* species (Cushion et al. 2004; Keely et al. 2004; Dei-Cas et al. 2006). The absence of genetic flow between *Pneumocystis* isolates from different mammal species (Mazars et al. 1994), associated with a marked host-species related phenotypic divergence (Aliouat-Denis et al. 2008), strengthen the taxonomic and conceptual consistency of the five *Pneumocystis* species described hitherto. They can be shortly characterized as follows:

Pneumocystis jirovecii (Frenkel 1999) was the sole species identified in humans. It was described for the first time in 1976 in the framework of the International Code of Zoological Nomenclature (ICZN) on the basis of apparent host species specificity and antigenic differences with rat-derived *Pneumocystis* (Frenkel 1976). This author described again the species in 1999, this time in the framework of the International Code of Botanical Nomenclature (ICBN) (Frenkel 1999). At the genetic level *P. jirovecii* clustered clearly with the other primate strains, and showed significant DNA sequence divergence with rodent-derived *Pneumocystis* (Dei-Cas et al. 2006; Aliouat-Denis et al. 2008). Consistently, *P. jirovecii* organisms revealed unable to infect SCID mice, a result that suggests the species has close host

species-specificity (Durand-Joly et al. 2002). At the ultrastructural level, *P. jirovecii* trophic forms present thick filopodia, similar to those observed in macaca-derived *Pneumocystis* (Durand-Joly et al. 2000) and in the species *P. oryctolagi*, from rabbits (Dei-Cas et al. 2006).

Pneumocystis carinii (Frenkel 1999) is the type species of the genus. It was first recognized as an independent taxonomic entity by Delanoë and Delanoë (1912) in *Rattus norvegicus*. Afterward, as steroid-treated rats were the most frequently used PcP models, most available data on *Pneumocystis* biology, culture, susceptibility to drugs, pathology, etc., were obtained from *P. carinii*. As a matter of fact, the genome of *P. carinii* is the first being sequenced in the framework of the "*Pneumocystis* Genome Project" (http://pneumocystis.cchmc.org/). Laboratory rats, either steroid-treated conventional rats or nude rats provide the highest parasite loads. Ultrastructurally, filopodia of *P. carinii* are thinner than those from rabbit or primate *Pneumocystis* species (see Dei-Cas et al. 2004, for review). The in vivo doubling time of *P. carinii* was found to be of 4.5 days in steroid-treated rats (Aliouat et al. 1999). Histologically, rat PcP evokes strongly human or primate PcP (Durand-Joly et al. 2000), and most therapeutic protocols against human PcP were based on results of experimental work performed in the rat model. *P. carinii* seems to display a stronger in vitro binding capacity to target cells than *P. murina* does. In vivo, *P. carinii* was unable to infect SCID mice (Aliouat et al. 1993b, 1994).

Pneumocystis wakefieldiae (Cushion et al. 2004) is the second species formally described in laboratory rats (*R. norvegicus*). The species was characterized on the basis of electrophoretic karyotype, localization of eight orthologous genes on different chromosomes, sequence divergence of several genes (4–7% divergence in seven orthologous genes between *P. wakefieldiae* and *P. carinii*), antigenic variation and MSG expression (Vasquez et al. 1996; Cushion 1998; Cushion et al. 2004; Schaffzin and Stringer 2000). A PCR method was developed to discriminate between *P. wakefieldiae* and *P. carinii* (Palmer et al. 1999). In mixed infections, the two species may also be distinguished using a single-strand conformation polymorphism method (Nahimana et al. 2001). Intriguingly, *P. wakefieldiae* is often found associated with *P. carinii* in laboratory or wild rats (Cushion et al. 2004; Cushion 1998). Nevertheless, in wild *R. norvegicus* from Denmark (Palmer et al. 2000) or Thailand (Chabé et al. 2010), *P. wakefieldiae* DNA was the sole amplified sequence in a significant number of cases. These observations associated to the genetic stability of this species in laboratory rats (Cushion et al. 1993) strengthened the validity of the taxon.

Pneumocystis murina (Keely et al. 2004) is the sole species identified in laboratory mice (*Mus musculus*). The species was recently described in the framework of ICBN rules (Keely et al. 2004) on the basis of PSC. It clustered clearly with the other rodent-derived *Pneumocystis* (Guillot et al. 2001) being close to rat-derived *Pneumocystis* species. Like in rat *Pneumocystis*, *P. murina* filopodia are more numerous, thin, and tree-like when compared to those of rabbit or primate *Pneumocystis* species (Dei-Cas et al. 1994, 2004, 2006; Nielsen et al. 1998; Durand-Joly et al. 2000). Additionally, *P. murina* trophic forms usually display more abundant cytoplasmic dense granules than those of the other species (Nielsen et al. 1998).

The doubling time of *P. murina* was found to be of 10.5 days in SCID mice (Aliouat et al. 1999). Organisms of this species were found to have weaker in vitro binding capacity to target cells when compared to *P. carinii* (Aliouat et al. 1993a). According to our own experience, latent *Pneumocystis* parasitism seemed to be less frequent in mice than in rats or rabbits. However, mice have been largely used both as parasite source and as experimental hosts for studies on *Pneumocystis* transmission, genetic polymorphism, host–parasite relationships, immune response, etc.

Pneumocystis oryctolagi (Dei-Cas et al. 2006) is the sole species described in Old World rabbits (*Oryctolagus cuniculus*). This species induces usually benign PcP in rabbits at weaning (about 1 month after birth) without corticosteroid administration (Soulez et al. 1989; Rajagopalan-Levasseur et al. 1998; Dei-Cas et al. 2006). For this reason, the rabbit model has been used for studying *Pneumocystis*–surfactant interactions (Prévost et al. 1997; Aliouat et al. 1998) and host immune response (Allaert et al. 1996, 1997; Rajagopalan-Levasseur et al. 1998; Tamburrini et al. 1999), two processes influenced by corticosteroid administration. At the light microscope level, *P. oryctolagi* organisms cannot be distinguished unequivocally from other *Pneumocystis* species. Furthermore, in rabbits with spontaneous PcP the cystic-to-trophic form ratio is usually higher (about 0.10–0.15) than in immunosuppressed rodents with PcP (about 0.02–0.05). Interestingly, *P. oryctolagi* in vivo doubling time (1.7 days) is much shorter than in rodent *Pneumocystis* species (Aliouat et al. 1999; Dei-Cas et al. 2006).

8.5 Understanding the Circulation of *Pneumocystis* in Host Populations Using Molecular Tools

Advances made over the last 20 years in the understanding of the natural history of *Pneumocystis* infection have been largely due to the use of molecular biologic approaches. Indeed, such studies have improved understanding of PcP epidemiology, shedding light on *Pneumocystis* reservoir, transmission patterns and circulation of *Pneumocystis* in host populations.

On the basis of the identification of several species in the genus and the demonstration that each *Pneumocystis* species can only infect its own specific host (Aliouat-Denis et al. 2008), emerged clearly the idea that pneumocystosis is not a zoonosis but an anthroponosis. Consequently, it is highly probable that only humans serve as reservoirs to *Pneumocystis jirovecii*, the sole species found in men (Dei-Cas et al. 1998; Wakefield 1998; Dei-Cas 2000). For a long time, it was thought that PcP resulted from the reactivation of *Pneumocystis* latent forms because of the high prevalence of *Pneumocystis* primary infection in small children, and because provoked steroid immunosuppression in laboratory animals led currently to PcP. But today, a lot of evidences suggest that *Pneumocystis* infection is usually exogenous, i.e. it results rather from *de novo* infection than from reactivation. For example, we know that new *Pneumocystis* strains, as identified by molecular typing, are usually responsible for recurrent episodes of PcP in a same patient (see Chabé et al. 2009, for review).

In addition, with the healing of PcP, *Pneumocystis* organisms would be radically eliminated from the lungs, according to negative results of highly sensitive PCR tests (see Chabé et al. 2009, for review).

In humans, beyond the well-known PcP of immunodepressed patients, especially of AIDS patients, the use of non-invasive sampling methods and highly efficient PCR assays revealed the presence of *Pneumocystis* DNA in a variable spectrum of health situations (Aliouat-Denis et al. 2008). For instance, *Pneumocystis* colonization or carriage (i.e. detection of low rates of *Pneumocystis* organisms or their DNA, without pneumonia) was described in healthy or hospitalized subjects without severe immunodepression like healthy small children, adults, pregnant women and subjects with chronic respiratory conditions, e.g. chronic obstructive pulmonary disease (COPD) (Calderon et al. 1996, 2004, 2007; Dei-Cas 2000; Peterson and Cushion 2005; Aliouat-Denis et al. 2008).

The role of immunocompetent colonized hosts in *Pneumocystis* epidemiology was investigated recently. Using the SCID–Balb/c mouse airborne transmission system described by Dumoulin et al. (2000), it was shown that healthy host-to-healthy host transmission of *P. murina* organisms can occur. More importantly, the 'second' healthy contacts revealed to be able to transmit the infectious organisms to immunocompromised, susceptible hosts (Gigliotti et al. 2003; Chabé et al. 2004). In addition, using histology and testing the expression of both cyclin-dependent serine-threonine kinase and heat-shock 70 protein in *Pneumocystis*, the authors showed that *Pneumocystis* organisms are able to dwell and replicate in the lungs of immunocompetent hosts, which points out these hosts as reservoir for *Pneumocystis* species (Chabé et al. 2004). Thus, healthy hosts behave actually as transient *Pneumocystis* carriers constituting a sort of *dynamic reservoir* as they are able to radically eliminate the parasites from their lungs, but as long as they remain infected they may transmit the infection either to naive hosts, or to immunosuppressed members of the population prone to develop PcP (Chabé et al. 2004). Like in mice, *P jirovecii* could circulate by airborne route in subpopulations more or less able to fight against the infection.

In order to approach the epidemiology of *Pneumocystis* infection, the main genotyping methods that have been used are the following: DNA sequence analysis, multitarget PCR-single-strand conformation polymorphism (SSCP), analysis of tandem repeats in the intron of the expression site of the *P. jirovecii* major surface glycoprotein (MSG), and more recently, a multilocus sequence typing method based on six distinct loci (Beard et al. 2004; Esteves et al. 2010).

Concerning the circulation of *P. jirovecii*, several molecular studies testified to the idea of interindividual airborne transmission in humans in both hospitals, as a nosocomial infection (Nevez et al. 2008), and community (see Chabé et al. 2009 for review). For example, studies using a multi-target PCR-SSCP method or *P. jirovecii* genotyping at the internal transcribed spacers (ITS) of the nuclear rRNA operon suggested that interhuman transmission of *P. jirovecii* occurred in PcP clusters among patients in hospitals (Nevez et al. 2008). *Pneumocystis* DNA was also detected in healthcare workers exposed to patients with PcP (Vargas et al. 2000). Moreover, investigations based on a genotyping approach of subjects colonized with *P. jirovecii* like small children developing *Pneumocystis* primary infection or

other immunocompetent *Pneumocystis* carriers, have suggested that these populations have a role in the circulation or transmission of the microorganism (Totet et al. 2004; Rivero et al. 2008). The analysis of dihydropteroate synthase (DHPS) locus of *P. jirovecii* was also useful to understand the circulation of the microorganism in the human reservoir. DHPS is the enzymatic target of sulfonamides, which are the major drugs for PcP prophylaxis and treatment. Now, prior exposure to sulfonamide drugs has been identified as a predictor of *P. jirovecii* DHPS mutant genotypes, which in other infections are usually associated with sulfa-drug resistance (reviewed in Totet et al. 2004). However, a high frequency of *P. jirovecii* DHPS mutations was recorded in patients who had never used sulfamide-based PcP prophylaxis or treatment (Huang et al. 2001; Kazanjian et al. 2000). Also, the city in which a patient resides has been identified as an independent risk factor of *Pneumocystis* DHPS mutation (Huang et al. 2001). On the whole, the evoked observations support the hypothesis that *P. jirovecii* can be transmitted from sulfa-treated patients to untreated subjects, potentially susceptible to *P. jirovecii*.

Airborne route could explain the transmission of the fungus between the members of human populations, but although the respiratory route is very likely the most common and important mode of transmission of *Pneumocystis*, the existence of other routes of transmission–like transplacental one- cannot be ruled out. Indeed, vertical transmission of *Pneumocystis* via the transplacental route was demonstrated in rabbits, but it does not appear to occur in rats or SCID mice (see Chabé et al. 2009 for review). In humans, congenital transmission was suspected for many years, and a recent report documented the presence of *Pneumocystis jirovecii* DNA in foetal lung and placenta samples, recovered from nonimmunodepressed pregnant women who had a miscarriage (Montes-Cano et al. 2009). The presence of *Pneumocystis* DNA in fetal lungs was also well documented in another primate (Demanche et al. 2003).

So, to summarize, *Pneumocystis jirovecii* could circulate actively by airborne route between the members of human populations, who present usually heterogeneous levels of innate or adaptive immune defense capacities. And we see that, in fact, PcP is a rare event in the natural history of *Pneumocystis* infection, and that beside *Pneumocystis* airborne transmission, a "protected" transplacental mother-to-offspring transmission could explain the high worldwide prevalence of *Pneumocystis* infection in small children.

To understand *Pneumocystis* circulation in ecosystems, animal patterns of transmission turn out to be crucial. Of particular interest was a study in non-human primates, where highly sensitive *Pneumocystis* molecular detection techniques associated to the use of efficient noninvasive sampling methods (Demanche et al. 2005) revealed a constant and intensive circulation of *Pneumocystis* organisms within a social organization of macaques (*Macaca fascicularis*) (Demanche et al. 2005).

Using mostly microscopic methods, the presence of *Pneumocystis* organisms was also reported in a large variety of zoo, domestic or wild mammals from different continents (Table 8.1). But compared to light microscopy, molecular methods have definite advantages for approaching *Pneumocystis* infection in wild hosts,

especially higher sensitivity and discriminatory power. Actually, molecular approaches allow identifying *Pneumocystis* species or strains, which are usually indistinguishable one another by using light microscopy and do not grow in in vitro culture. The molecular characterization of *Pneumocystis* species dwelling in the lungs of wild animals remains therefore a crucial way to understand how these pathogens disseminate in ecosystems.

On the whole, the exploring of *Pneumocystis* in wild or domestic mammals by using molecular techniques showed typically the following facts: (i) a high prevalence of *Pneumocystis* colonization is found in wild mammals, probably resulting from active airborne horizontal and vertical (transplacental or aerial) transmission mechanisms; (ii) a specific *Pneumocystis* DNA sequence could be attributed to each host species; (iii) *P. jirovecii* DNA was never detected in the lungs of non-human mammals; (iv) although fatal *Pneumocystis* infections were recorded in newly captured coatimundis, *Nasua narica*, and sloths, *Bradypus tridactylus* and *Bradypus variegatus* (Edentata) in Brasil and Panama (Lainson and Shaw 1975; Yonushonis et al. 1986), PcP cases were reported rarely in wild mammal populations living in natural conditions (Aliouat-Denis et al. 2008, Table 8.1). The infection occurs in apparently immunodepressed animals but the type and the degree of immunodeficiency could not be established in most cases. Thus, in natural ecosystems *Pneumocystis* organisms circulate through the members of host populations developing mostly mild parasitism, though frequent, in the lungs of immunocompetent hosts.

In conclusion, *Pneumocystis* organisms illustrate the concept of phylogenetic specificity exhibited by groups of highly adapted parasites showing low pathogenicity, high host specificity maintained throughout time, extensive colonization of host populations, and a life cycle closely linked with host organs and physiology (Humphery-Smith 1989).

8.6 How *Pneumocystis* Molecular Phylogeny May Improve the Understanding of Host Taxonomy

Several authors have suggested that the phylogenetic relationships of highly host specific parasites would provide valuable information about the evolutionary history of their hosts (see Brooks and McLennan 1993 for a review). Sometimes, the life-histories of two different lineages are so intimately linked that a speciation in one group induces a parallel speciation event in the other. If cophylogeny was the only process occurring, the host and parasite phylogenies would exactly mirror each other making it possible to base genealogical conclusions on parasite data. Parallel coevolution between the primates and their specific parasites yet was mentioned concerning parasites as different as viruses (Siddall 1997), nematodes (Hugot 1999), mites (O'Connor 1985) or lice (Reed et al. 2004). Comparison of the *Pneumocystis* phylogeny with the phylogeny of their primate hosts gives an additional example of cophylogeny (Hugot et al. 2003).

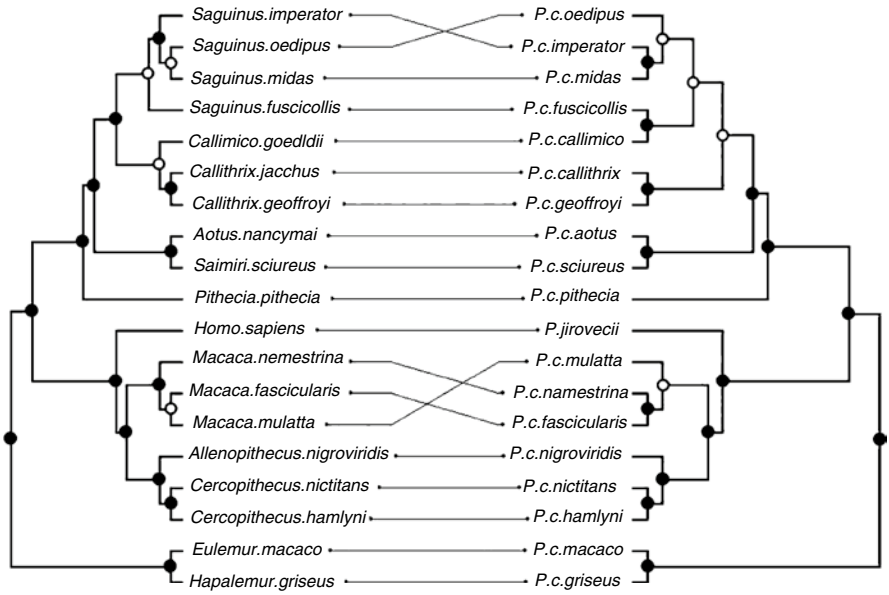

Fig. 8.2 Coevolution in Primate's *Pneumocystis* (after Hugot et al. 2003). The host tree (*on the left*) is after Fabre et al. (2009). The parasite tree (*on the right*) was built on the basis of three-concatenated gene polymorphism (mtLSU rDNA, mtSSU rDNA, DHPS). *Black spots* and *open spots* showed congruent and incongruent nodes, respectively (*see text*)

Figure 8.2 allows comparing the phylogenetic tree of the *Pneumocystis* parasite species with a phylogeny of the corresponding primate host species. The parasite phylogeny is after Hugot et al. (2003), the tree resulting of the ML analysis of a data set, combining three molecular markers (DHPS, dehydropteroate synthetase; mtSSU-rRNA, small subunit of mitochondrial ribosomal RNA; mtLSU-rRNA, large subunit of mitochondrial ribosomal RNA). The host tree is after Fabre et al. (2009). Comparison of the host and parasite trees was performed using TreeMap (Page 1995): an algorithm to find all reconstructions that maximize the number of codivergences in the particular case of a host-parasite assemblage.

Congruent (corresponding) nodes are underlined using black spots and incongruent nodes using open spots. Only four apparent incongruences emerge between parasite tree *versus* host tree. There are at the level of *Saguinus* species, *Callimico* and *Callithrix* genera and of *Macaca* species (Fig. 8.2). More than 77% of the nodes (14 out of 18) are congruent and the probability to obtain such a result by chance is very low. In addition, when this work was first published, the relationships between *Pithecia*, *Aotus* and *Saimiri* were not definitively resolved; since that time progress have been performed and the phylogenetic relationships between these three genera (Fabre et al. 2009), revealed to be identical to the phylogeny of the corresponding *Pneumocystis* species. Thus, in this case the phylogeny of the parasites enlightened the phylogeny of the hosts.

8.7 Conclusion

On the whole, molecular approaches used in *Pneumocystis* research have played a crucial role in the shift from an old conceptual *Pneumocystis* infection framework to a new one. Particularly, molecular methodologies, associated with other approaches, have drastically changed our views on the epidemiology and natural history of pneumocystosis. In the old conceptual framework, *Pneumocystis* organisms were considered undefined or even enigmatic protists belonging to a unique euryxenic taxonomic entity ("*P. carinii*") transmissible by the airborne route between mammals of different species. In the new conceptual framework, the *Pneumocystis* genus is a highly diversified group of parasitic microfungi that contains numerous stenoxenic species closely adapted to, and coevolving with, mammal species. Thus, present developments in molecular phylogeny are leading us to identify new *Pneumocystis* species on the basis of Phylogenetic Species Concept. Finally, thanks to the high host-specificity of *Pneumocystis* organisms and their long-range adaptation to mammals, *Pneumocystis* species and genotypes could be useful tools to understand taxonomy and phylogeny of their mammal hosts. In other words, the topology of *Pneumocystis* phylogenetic trees could shed light on the mammalian host evolution history, helping to solve taxonomic uncertainties.

References

Aliouat EM, Dei-Cas E, Ouaissi A et al (1993a) In vitro attachment of *Pneumocystis carinii* from mouse and rat origin. Biol Cell 77:209–217

Aliouat EM, Mazars E, Dei-Cas E et al (1993b) Intranasal inoculation of mouse, rat or rabbit-derived *Pneumocystis* in SCID mice. J Protozool Res 3:94–98

Aliouat EM, Mazars E, Dei-Cas E et al (1994) *Pneumocystis* cross infection experiments using SCID mice and nude rats as recipient host, showed strong host-species specificity. J Eukaryot Microbiol 41:S71

Aliouat EM, Escamilla R, Cariven C et al (1998) Surfactant changes during experimental pneumocystosis are related to the *Pneumocystis* development. Eur Respir J 11:542–547

Aliouat EM, Dujardin L, Martinez A et al (1999) *Pneumocystis carinii* growth kinetics in culture systems and in hosts: involvement of each life cycle parasite stage. J Eukaryot Microbiol 46:S116–S117

Aliouat-Denis CM, Chabé M, Demanche C et al (2008) *Pneumocystis* species, co-evolution and pathogenic power. Infect Genet Evol 8:708–726

Aliouat-Denis CM, Martinez A, Aliouat EM et al (2009) The *Pneumocystis* life cycle. Mem Inst Oswaldo Cruz 104:419–426

Allaert A, Levasseur P, Jouault T et al (1996) Role of alveolar macrophages during the spontaneous *Pneumocystis carinii* pneumonia of rabbits at weaning. J Eukaryot Microbiol 43:23S

Allaert A, Jouault T, Rajagopalan-Levasseur P et al (1997) Detection of cytokine mRNA in the lung during spontaneous *Pneumocystis carinii* pneumonia of the young rabbit. J Eukaryot Microbiol 44:45S

Banerji S, Lugli EB, Miller RF et al (1995) Analysis of genetic diversity at the arom locus in isolates of *Pneumocystis carinii*. J Eukaryot Microbiol 42:675–679

Barton EG, Campbell WG (1967) Further observations on the ultrastructure of *Pneumocystis*. Arch Pathol 83:527–534

Bauer NL, Paulsrud JR, Bartlett MS et al (1993) *Pneumocystis carinii* organisms obtained from rats, ferrets and mice are antigenically different. Infect Immun 61:1315–1319
Beard CB, Roux P, Nevez G et al (2004) Strain typing methods and molecular epidemiology of *Pneumocystis* pneumonia. Emerg Infect Dis 10:1729–1735
Beck JM, Preston AM, Wagner JG et al (1998) Interaction of rat *Pneumocystis carinii* and rat alveolar epithelial cells *in vitro*. Am J Physiol 275:L118–L125
Bishop R, Gurnell J, Laakkonen J et al (1997) Detection of *Pneumocystis* DNA in the lungs of several species of wild mammal. J Eukaryot Microbiol 44:57S
Blazec K (1960) Die *Pneumocystis* – pneumonia beim feldhasen (Lepus europeus Pallas, 1778). Zentralbl Allg Pathol Anat 101:484–489
Brooks DR, McLennan DA (1993) Parascript. In: Funk VA, Cannell PF (eds) Parasites and the language of evolution. Smithsonian Institution Press, Washington, DC, p 429
Burgess JW, Kottom TJ, Limper AH (2008) *Pneumocystis carinii* exhibits a conserved meiotic control pathway. Infect Immun 76:417–425
Cailliez JC, Séguy N, Denis CM et al (1996) *Pneumocystis carinii*: an atypical fungal microorganism. J Med Vet Mycol 34:227–239
Calderón EJ, Regordan C, Medrano FJ et al (1996) *Pneumocystis carinii* infection in patients with chronic bronchial disease. Lancet 347:977
Calderón EJ, Varela JM, Medrano FJ et al (2004) Epidemiology of *Pneumocystis carinii* pneumonia in southern Spain. Clin Microbiol Infect 10:673–676
Calderón EJ, Rivero L, Respaldiza N et al (2007) Systemic inflammation in patients with chronic obstructive pulmonary disease who are colonized with *Pneumocystis jirovecii*. Clin Infect Dis 45:17–19
Campbell WG (1972) Ultrastructure of *Pneumocystis* in human lungs: life cycle of human pneumocystosis. Arch Pathol 93:312–324
Cavallini Sanches EM, Pacheco SM, Cericatto AS et al (2009) Detection of *Pneumocystis* in lungs of bats from Brazil by PCR amplification. Pesqui Vet Bras 29:469–473
Ceré N, Drouet-Viard F, Dei-Cas E et al (1997) *In utero* transmission of *Pneumocystis carinii* sp. f. *oryctolagi*. Parasite 4:325–330
Chabé M, Dei-Cas E, Creusy C et al (2004) Immunocompetent hosts as a reservoir of *Pneumocystis* organisms: histological and RT-PCR data demonstrate active replication. Eur J Clin Microbiol Infect Dis 23:89–97
Chabé M, Nevez G, Totet A et al (2009) Transmission de *Pneumocystis*. J Mycol Med 19:276–284
Chabé M, Herbreteau V, Hugot JP et al (2010) *Pneumocystis carinii* and *Pneumocystis wakefieldiae* in wild *Rattus norvegicus* trapped in Thailand. J Eukaryot Microbiol 57:213–217
Chabé M, Aliouat-Denis CM, Delhaes L et al (2011) *Pneumocystis*: from a doubtful unique entity to a group of highly diversified fungal species. FEMS Yeast Res 11:2–17
Chagas C (1909) Nova tripanozomiazaea humana. Mem Inst Oswaldo Cruz 1:159–218
Christensen CBV, Settnes OP, Bille-Hansen V et al (1996) *Pneumocystis carinii* from pigs and humans are antigenically distinct. J Med Vet Mycol 34:431–433
Cornillot E, Keller B, Cushion MT et al (2002) Fine analysis of the *Pneumocystis carinii* f. sp. *carinii* genome by two-dimensional pulsed-field gel electrophoresis. Gene 293:87–95
Creusy C, Bahon-le Capon J, Fleurisse L et al (1996) *Pneumocystis carinii* pneumonia in four mammal species: histopathology and ultrastructure. J Eukaryot Microbiol 43:47S–48S
Cross JH, Hung CK (1975) Latent *P. carinii* infections in some mammals from Taipei Hsien, Taiwan. J Formos Med Assoc 74:564–567
Cushion MT (1998) Genetic heterogeneity of rat-derived *Pneumocystis*. FEMS Immunol Med Microbiol 22:51–58
Cushion MT (2004) *Pneumocystis*: unraveling the cloak of obscurity. Trends Microbiol 12:243–249
Cushion MT, Kaselis M, Stringer SL et al (1993) Genetic stability and diversity of *Pneumocystis carinii* infecting rat colonies. Infect Immun 61:4801–4813
Cushion M, Keely S, Stringer J (2004) Molecular and phenotypic description of *Pneumocystis wakefieldiae* sp.nov., a new species in rats. Mycologia 96:429–438

Cushion MT, Smulian AG, Slaven BE et al (2007) Transcriptome of *Pneumocystis carinii* during fulminate infection: carbohydrate metabolism and the concept of a compatible parasite. PLoS One 2:e423. doi:10.1371/journal.pone.0000423

Cushion MT, Linke MJ, Ashbaugh A et al (2010) Echinocandin treatment of *Pneumocystis* pneumonia in rodent models depletes cysts leaving trophic burdens that cannot transmit the infection. PLoS One 5:e8524

De Hoog GS, Guarro J, Gené J et al (eds) (2000) Atlas of clinical fungi, 2nd edn. Centraalbureau voor Schimmelcultures (CBS)/Universitat Rovira i Virgili, Utrecht/Reus

de Souza W, Benchimol M (2005) Basic biology of *Pneumocystis carinii*: a mini review. Mem Inst Oswaldo Cruz 100:903–908

De Stefano JA, Cushion MT, Sleught R et al (1990) Analysis of *Pneumocystis carinii* cyst wall. I. Evidence for an outer surface membrane. J Protozool 37:428–435

Dei-Cas E (2000) *Pneumocystis* Infections: the iceberg? Med Mycol 1:23–32

Dei-Cas E, Vernes A (1986) Parasitic adaptation of pathogenic fungi to the mammalian hosts. CRC Crit Rev Microbiol 13:173–218

Dei-Cas E, Jackson H, Palluault F et al (1991) Ultrastructural observations on the attachment of *Pneumocystis carinii in vitro*. J Protozool 38:S205–S207

Dei-Cas E, Cailliez JC, Palluault F et al (1992) Is *Pneumocystis carinii* a deep mycosis-like agent? Eur J Epidemiol 8:460–470

Dei-Cas E, Mazars E, Ödberg-Ferragut C et al (1994) Ultrastructural, genomic, isoenzymatic and biological features make it possible to distinguish rabbit *Pneumocystis* from other mammal *Pneumocystis* strains. J Eukaryot Microbiol 41:S84

Dei-Cas E, Mazars E, Aliouat EM et al (1998) The host-specificity of *Pneumocystis carinii*. J Mycol Med 8:1–6

Dei-Cas E, Aliouat EM, Cailliez JC (2004) Cellular structure. In: Walzer PD, Cushion MT (eds) *Pneumocystis* pneumonia, 3rd edn. Marcel Dekker Inc, New York, pp 61–94

Dei-Cas E, Chabé M, Moukhlis R et al (2006) *Pneumocystis oryctolagi* sp. nov., an uncultured fungus causing pneumonia in rabbits at weaning: review of current knowledge, and description of a new taxon on genotypic, phylogenetic and phenotypic bases. FEMS Microbiol Rev 80:853–871

Delanoë P, Delanoë E (1912) Sur les rapports des kystes de carinii du poumon des rats avec le *Trypanosoma lewisi*. CR Acad Sci 155:658–660

Demanche C, Berthelemy M, Petit T et al (2001) Phylogeny of *Pneumocystis carinii* from 18 primate species confirms host specificity and suggests coevolution. J Clin Microbiol 39:2126–2133

Demanche C, Petit T, Moisson P et al (2003) Assessment of *Pneumocystis* species carriage in captive primates. Vet Rec 152:811–813

Demanche C, Wanert F, Barthelemy M et al (2005) Molecular and serological evidence of *Pneumocystis* circulation in a social organization of healthy macaques (*Macaca fascicularis*). Microbiology 151:3117–3125

Denis CM, Mazars E, Guyot K et al (2000) Genetic divergence at the SOD-A locus of six different formae speciales of *Pneumocystis carinii*. Med Mycol 38:289–300

Derouiche S, Deville M, Taylor ML et al (2009) *Pneumocystis* diversity as a phylogeographic tool. Mem Inst Oswaldo Cruz 104:112–117

Doby JM, Kombila-Favry M (1978) Presence of sexual forms (cleistothecia and Hülle cells) in a case of human maxillary sinus aspergillosis caused by Aspergillus nidulans in association with Aspergillus fumigatus. Mycopathologia 64:157–163

Dumoulin A, Mazars E, Seguy N et al (2000) Transmission of *Pneumocystis carinii* disease from immunocompetent contacts of infected hosts to susceptible hosts. Eur J Clin Microbiol Infect Dis 19:671–678

Durand-Joly I, Wakefield AE, Palmer RJ et al (2000) Ultrastructural and molecular characterization of *Pneumocystis carinii* isolated from a rhesus monkey (*Macaca mulatta*). Med Mycol 38:61–72

Durand-Joly I, Aliouat EM, Recourt C et al (2002) *Pneumocystis carinii* f. sp. *hominis* is not infectious for SCID mice. J Clin Microbiol 40:1862–1865

Edman JC, Kovacs JA, Masur H et al (1988) Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. Nature 334:519–522

Edman JC, Kovacs JA, Masur H et al (1989) Ribosomal RNA sequence shows *Pneumocystis carinii* to be a member of the fungi. J Protozool 36:S18–S20

English K, Peters SE, Maskell DJ et al (2001) DNA analysis of *Pneumocystis* infecting a Cavalier King Charles spaniel. J Eukaryot Microbiol 48:106S

Eriksson OE (1994) *Pneumocystis carinii*, a parasite in lungs of mammals, referred to a new family and order (*Pneumocystidaceae, Pneumocystidales, Ascomycota*). Systema Ascomycetum 13:165–180

Eriksson OE, Winka K (1997) Supraordinal taxa of the Ascomycota. Myconet 1:1–16

Esteves F, Gaspar J, Tavares A et al (2010) Population structure of *Pneumocystis jirovecii* isolated from immunodeficiency virus-positive patients. Infect Genet Evol 10:192–199

Fabre PH, Rodrigues A, Douzery EJP (2009) Patterns of macroevolution among primates inferred from a supermatrix of mitochondrial and nuclear DNA. Mol Phylogenet Evol 53:808–825

Fischer JM, Keely SP, Stringer JR (2006) Evolutionary rate of Ribosomal DNA in *Pneumocystis* species is normal despite the extraordinarily low copy-number of rDNA genes. J Eukaryot Microbiol 53:S156–S158

François F, Noël T, Pépin R et al (2001) Alternative identification test relying upon sexual reproductive abilities of Candida lusitaniae strains isolated from hospitalized patients. J Clin Microbiol 39:3906–3914

Fréalle E, Noël C, Viscogliosi E et al (2005) Manganese superoxide dismutase in pathogenic fungi: an issue with pathophysiological and phylogenetic involvements. FEMS Immunol Med Microbiol 45:411–422

Fréalle E, Noël C, Nolard N et al (2006) Manganese superoxide dismutase based phylogeny of pathogenic fungi. Mol Phylogenet Evol 41:28–39

Frenkel JK (1976) *Pneumocystis jiroveci* n. sp. from man: morphology, physiology, and immunology in relation to pathology. Natl Cancer Inst Monogr 43:13–30

Frenkel JK (1999) *Pneumocystis* pneumonia, an immunodeficiency-dependent disease (IDD): a critical historical overview. J Eukaryot Microbiol 46:S89–S92

Furuta T, Ueda K (1987) Intra- and interspecies transmission and antigenic difference of *Pneumocystis carinii* derived from rat and mouse. Jpn J Exp Med 57:11–17

Gigliotti F (1992) Host species-specific antigenic variation of a mannosylated surface glycoprotein of *Pneumocystis carinii*. J Infect Dis 165:329–336

Gigliotti F, Harmsen AG, Haidaris CG et al (1993) *Pneumocystis carinii* is not universally transmissible between mammalian species. Infect Immun 61:2886–2890

Gigliotti F, Harmsen AG, Wright TW (2003) Characterization of transmission of *Pneumocystis carinii* f.sp. *muris* through immunocompetent BALB/c mice. Infect Immun 71:3852–3856

Giuntoli D, Stringer SL, Stringer JR (1994) Extraordinarily low number of ribosomal RNA genes in *P. carinii*. J Eukaryot Microbiol 41:88S

Goyot P, Mojon M, Petavy AF (1986) *Arvicola terrestris*, Linnaeus 1758, a reservoir of *Pneumocystis* sp. Delanoë and Delanoë 1912. Z Parasitenkd 72:685–687

Guarro J, Gené J, Stchigel AM (1999) Developments in fungal taxonomy. Clin Microbiol Rev 12:454–500

Guillot J, Chevalier V, Queney G et al (1999) Acquisition and biodiversity of *Pneumocystis carinii* in a colony of wild rabbits (*Oryctolagus cuniculus*). J Eukaryot Microbiol 46:100S–101S

Guillot J, Demanche C, Hugot JP et al (2001) Parallel phylogenies of *Pneumocystis* species and their mammalian hosts. J Eukaryot Microbiol 47:S113–S115

Haque AU, Plattner SB, Cook RT et al (1987) *Pneumocystis carinii*: taxonomy as viewed by electron microscopy. Am J Clin Pathol 87:504–510

Huang L, Friedly J, Morris AM et al (2001) *Pneumocystis carinii* dihydropteroate synthase genotypes in HIV-infected persons residing in San Francisco: possible implications for disease transmission. J Eukaryot Microbiol 48:137S–138S

Hugot JP (1999) Primates and their pinworm parasites: the Cameron hypothesis revisited. Syst Biol 48:523–546

Hugot JP, Demanche C, Barriel V et al (2003) Phylogenetic systematics and evolution of primate-derived *Pneumocystis* based on mitochondrial or nuclear DNA sequence comparison. Syst Biol 52:735–744

Humphery-Smith I (1989) The evolution of phylogenetic specificity among parasitic organisms. Parasitol Today 5:385–387

Hunter JA, Wakefield AE (1996) Genetic divergence at the mitochondrial small subunit ribosomal RNA gene among isolates of *Pneumocystis carinii* from five mammalian host species. J Eukaryot Microbiol 43:S24–S25

Itatani CA (1994) Ultrastructural demonstration of a pore in the cyst wall of *Pneumocystis carinii*. J Parasitol 80:644–648

Itatani CA (1996) Ultrastructural morphology of intermediate forms and forms suggestive of conjugation in the life cycle of *Pneumocystis carinii*. J Parasitol 82:163–171

Ito M, Nozu R, Kuramochi T et al (2000) Prophylactic effect of FK463, a novel antifungal lipopeptide, against *Pneumocystis carinii* infection in mice. Antimicrob Agents Chemother 44:2259–2262

Kaneshiro ES, Collins MS, Cushion MT (2000) Inhibitors of sterol biosynthesis and amphotericin B reduce the viability of *Pneumocystis carinii* f.sp. *carinii*. Antimicrob Agents Chemother 44:1630–1638

Kazanjian P, Armstrong W, Hossler PA et al (2000) *Pneumocystis carinii* mutations are associated with duration of sulfa or sulfone prophylaxis exposure in AIDS patients. J Infect Dis 182:551–557

Keely SP, Fischer JM, Stringer JR (2003) Evolution and speciation of *Pneumocystis*. J Eukaryot Microbiol 50:S624–S626

Keely SP, Fischer JM, Cushion MT et al (2004) Phylogenetic identification of *Pneumocystis murina* sp. nov., a new species in laboratory mice. Microbiology 150:1153–1165

Kirkbride CA (1992) Etiologic agents detected in a 10-year study of bovine abortions and stillbirths. J Vet Diagn Invest 4:175–180

Kondo H, Taguchi M, Abe N et al (1993) Pathological changes in epidemic porcine *Pneumocystis carinii* pneumonia. J Comp Pathol 108:261–268

Kottom TJ, Thomas CF Jr, Mubarak KK et al (2000) *Pneumocystis carinii* uses a functional cdc13 B-type cyclin complex during its life cycle. Am J Respir Cell Mol Biol 22:722–731

Kottom TJ, Köhler JR, Thomas CF et al (2003) Lung epithelial cells and extracellular matrix components induce expression of *Pneumocystis carinii* STE20, a gene complementing the mating and pseudohyphal growth defects of STE20 mutant yeast. Infect Immun 71:6463–6471

Laakkonen J (1998) *Pneumocystis carinii* in wildlife. Int J Parasitol 28:241–252

Laakkonen J, Sukura A (1997) *Pneumocystis carinii* of the common shrew, *Sorex araneus*, shows a discrete phenotype. J Eukaryot Microbiol 44:117–121

Laakkonen J, Sukura A, Haukisalmi V et al (1993) *Pneumocystis carinii* and helminth parasitism in shrews *Sorex araneus* and *Sorex caecutiens*. J Wildl Dis 29:273–277

Laakkonen J, Fisher RN, Case TJ (2001) Pneumocystosis in wild small mammals from California. J Wildl Dis 37:408–412

Laakkonen J, Nyyssönen T, Hiltunen M et al (2006) Effects of *Protostrongylus* sp. and *Pneumocystis* sp. on the pulmonary tissue and the condition of mountain and brown hares from Finland. J Wildl Dis 42:780–787

Lainson R, Shaw JJ (1975) *Pneumocystis* and *Histoplasma* infections in wild animals from the Amazon region of Brazil. Trans R Soc Trop Med Hyg 69:505–508

Li Z, Feng X, Lu S et al (2008) Molecular phylogeny of *Pneumocystis* based on 5.8S rRNA gene and the internal transcribed spacers of rRNA gene sequences. Sci China C Life Sci 51: 445–452

Linke MJ, Cushion MT, Walzer PD (1989) Properties of the major antigens of rat and human *Pneumocystis carinii*. Infect Immun 57:1547–1555

Liu Y, Leigh JW, Brinkmann H et al (2009) Phylogenomic analyses support the monophyly of Taphrinomycotina, including Schizosaccharomyces fission yeasts. Mol Biol Evol 26:27–34

Lobetti RG, Leisewitz AL, Spencer JA (1996) *Pneumocystis carinii* in the miniature dachshund: case report and literature review. J Small Anim Pract 37:280–285

Ma L, Imamichi H, Sukura A et al (2001) Genetic divergence of the dihydrofolate reductase and dihydropteroate synthase genes in *Pneumocystis carinii* from 7 different host species. J Infect Dis 184:1485–1488

Matsumoto Y, Yoshida Y (1984) Sporogony in *Pneumocystis carinii*: synaptonemal complexes and meiotic nuclear divisions observed in precysts. J Protozool 31:420–428

Mazars E, Dei-Cas E (1998) Epidemiological and taxonomic impact of *Pneumocystis* biodiversity. FEMS Immunol Med Microbiol 22:75–80

Mazars E, Ödberg-Ferragut C, Durand I et al (1994) Genomic and isoenzymatic markers of *Pneumocystis* from different host species. J Eukaryot Microbiol 41:S104

Mazars E, Ödberg-Ferragut C, Dei-Cas E et al (1995) Polymorphism of *Pneumocystis carinii* from different host species. J Eukaryot Microbiol 42:26–32

Mazars E, Guyot K, Fourmaintraux S et al (1997a) Detection of *Pneumocystis* in European wild Animals. J Eukaryot Microbiol 44:39S

Mazars E, Guyot K, Durand I et al (1997b) Isoenzyme diversity in *Pneumocystis carinii* from rats, mice and rabbits. J Infect Dis 175:655–660

Mc Connell EE, Basson PA, Pienaar JG (1971) Pneumocystosis in a domestic goat. Onderstepoort J Vet Res 38:117–124

Mitchell RG, Chaplin AJ, Mackenzie DW (1987) *Emericella nidulans* in a maxillary sinus fungal mass. J Med Vet Mycol 25:339–341

Montes-Cano MA, Chabé M, Fontillon-Alberdi M et al (2009) Vertical transmission of *Pneumocystis jirovecii* in humans. Emerg Infect Dis 15:125–127

Nahimana A, Cushion MT, Blanc DS et al (2001) Rapid PCR-single-strand conformation polymorphism method to differentiate and estimate relative abundance of *Pneumocystis carinii* special forms infecting rats. J Clin Microbiol 39:4563–4565

Nevez G, Chabé M, Rabodonirina M et al (2008) Nosocomial *Pneumocystis jirovecii* infections. Parasite 15:359–365

Nielsen MH, Settnes OP, Aliouat EM et al (1998) Different ultrastructural morphology of *Pneumocystis carinii* derived from mice, rats, and rabbits. APMIS 106:771–779

Nishida H, Sugiyama J (1993) Phylogenetic relationships among *Taphrina, Saitoella*, and other higher fungi. Mol Biol Evol 10:431–436

O'Connell RJ, Panstruga R (2006) Tête à tête inside a plant cell: establishing compatibility between plants and biotrophic fungi and oomycetes. New Phytol 4:699–718

O'Connor BM (1985) Host associations and coevolutionary relationships of astigmatid mite parasites of new world primates. I. Families Psoroptidae and Audycoptidae. Fieldiana, Zoology (new series) 39:245–260

Ortiz-Rivera M, Liu Y, Felder R et al (1995) Comparison of coding and spacer region sequences of chromosomal rRNA-coding genes of two sequevars of *Pneumocystis carinii*. J Eukaryot Microbiol 42:44–49

Page RDM (1995) Parallel phylogenies: reconstructing the history of host-parasite assemblages. Cladistics 10:155–173

Palluault F, Slomianny C, Soulez B et al (1992) High osmotic pressure enables fine ultrastructural and cytochemical studies on *Pneumocystis carinii.* I: Epon embedding. Parasitol Res 78:437–444

Palmer RJ, Cushion MT, Wakefield AE (1999) Discrimination of rat-derived *Pneumocystis carinii* f. sp. *carinii* and *Pneumocystis carinii* f. sp. *ratti* using the polymerase chain reaction. Mol Cell Probes 13:147–155

Palmer RJ, Settnes OP, Lodal J et al (2000) Population structure of rat-derived *Pneumocystis carinii* in Danish wild rats. Appl Environ Microbiol 66:4954–4961

Perron Lepage MF, Gerber V, Suter MM (1999) A case of interstitial pneumonia associated with *Pneumocystis carinii* in a foal. Vet Pathol 36:621–624

Peters SE, English K, Laakkonen J et al (1994a) DNA analysis of *Pneumocystis carinii* infecting Finnish and English shrews. J Eukaryot Microbiol 41:108S

Peters SE, Wakefield AE, Whitwell KE et al (1994b) *Pneumocystis carinii* pneumonia in thoroughbred foals: identification of a genetically distinct organism by DNA amplification. J Clin Microbiol 32:213–216

Peters SE, English K, Rana A et al (2001) Synaptonemal complexes in the pre-cyst of *Pneumocystis carinii*. J Eukaryot Microbiol 48:S134

Peterson JC, Cushion MT (2005) *Pneumocystis*: not just pneumonia. Curr Opin Microbiol 8:393–398

Poelma FG (1975) Pneumocystis carinii in zoo animals. Parasitenkd 46:61

Powles MA, Liberator P, Anderson J et al (1998) Efficacy of MK-991 (L-743,872), a semisynthetic pneumocandin, in murine models of *Pneumocystis carinii*. Antimicrob Agents Chemother 42:1985–1989

Prévost MC, Aliouat EM, Escamilla R et al (1997) Pneumocystosis in humans or in corticosteroid-untreated animal models: interactions between pulmonary surfactant changes and *Pneumocystis carinii in vivo* or *in vitro* growth. J Eukaryot Microbiol 44:58S

Rajagopalan-Levasseur P, Allaert A, Dridba M et al (1998) Response to *Pneumocystis* infection in an immunocompetent host. FEMS Immunol Med Microbiol 22:107–121

Reed C, O'Donoghue JL (1979) A new guinea pig mutant with abnormal hair production and immunodeficiency. Lab Anim Sci 29:744–748

Reed DL, Smith VS, Hammond SL et al (2004) Genetic analysis of lice supports direct contact between modern and archaic humans. PLoS Biol 2:e340

Rivero L, de la Horra C, Montes-Cano MA et al (2008) *Pneumocystis jirovecii* transmission from immunocompetent carriers to infant. Emerg Infect Dis 14:1116–1118

Sanches EM, Pescador C, Rozza D et al (2007) Detection of *Pneumocystis* spp. in lung samples from pigs in Brazil. Med Mycol 45:395–399

Sanchez CA, Chabé M, Aliouat EM et al (2007) Exploring transplacental transmission of *Pneumocystis oryctolagi* in first-time pregnant and multiparous rabbit does. Med Mycol 45:701–707

Schaffzin JK, Stringer JR (2000) The major surface glycoprotein expression sites of two special forms of rat *Pneumocystis carinii* differ in structure. J Infect Dis 181:1729–1739

Settnes OP, Hasselager E (1984) Occurrence of *Pneumocystis carinii* Delanoë & Delanoë, 1912 in dogs and cats in Denmark. Nord Vet Med 36:179–181

Settnes OP, Henriksen SA (1989) *Pneumocystis carinii* in large domestic animals in Denmark. A preliminary report. Acta Vet Scand 30:437–440

Settnes OP, Lodal J (1980) Prevalence of *Pneumocystis carinii* Delanoë & Delanoë, 1912 in rodents in Denmark. Nord Vet Med 32:17–27

Settnes OP, Nielsen MJ (1991) Host-parasite relationship in *Pneumocystis carinii* infection: activation of the plasmalemmal vesicular system in type I alveolar epithelial cells. J Protozool 38:S174–S176

Settnes OP, Elvestad K, Clausen B (1986) *Pneumocystis carinii* Delanoë & Delanoë, 1912 found in lungs of freeliving animals in Denmark at autopsy. Nord Vet Med 38:11–15

Settnes OP, Bille-Hansen V, Jorsal SE, Henriksen SA (1991) The piglet as a potential model of Pneumocystis carinii pneumonia. J Protozool 38:140S–141S

Settnes OP, Nielsen PB, Bucala R et al (1994) A survey of birds in Denmark for the presence of *Pneumocystis carinii*. Avian Dis 38:1–10

Sexton AC, Howlett BJ (2006) Parallels in fungal pathogenesis on plant and animal hosts. Eukaryot Cell 5:1941–1949

Shimizu A, Kimura F, Kimura S (1985) Occurrence of *Pneumocystis carinii* in animals in Japan. Nippon Juigaku Zasshi 47:309–311

Shiota T, Shimada Y, Kurimoto H, Oikawa H (1990) *Pneumocystis carinii* infection in corticosteroid-treated cats. J Parasitol 76:441–445

Siddall ME (1997) The AIDS pandemic is new, but is HIV *not* new? Cladistics 13:267–274

Sinclair K, Wakefield AE, Banerji S et al (1991) *Pneumocystis carinii* organisms derived from rat and human hosts are genetically distinct. Mol Biochem Parasitol 45:183–184

Smulian AG, Sesterhenn T, Tanaka R et al (2001) The ste3 pheromone receptor gene of *Pneumocystis carinii* is surrounded by a cluster of signal transduction genes. Genetics 157:991–1002

Soulez B, Dei-Cas E, Charet P et al (1989) The young rabbit: a nonimmunosuppressed model for *Pneumocystis carinii* pneumonia. J Infect Dis 160:355–356

Stedman TT, Butler DR, Buck GA (1998) The HSP70 gene family in *Pneumocystis carinii*: molecular and phylogenetic characterization of cytoplasmic members. J Eukaryot Microbiol 45:589–599

Stringer JR, Stringer SL, Zhang J (1993) Molecular genetic distinction of *Pneumocystis carinii* from rats and humans. J Eukaryot Microbiol 40:733–741

Stringer JR, Cushion MT, Wakefield AE (2001) New nomenclature for the genus *Pneumocystis*. J Eukaryot Microbiol 48:184–189

Sukura A, Saari S, Järvinen AK, Olsson M, Kärkkäinen M, Ilvesniemi T (1996) *Pneumocystis carinii pneumonia* in dogs–a diagnostic challenge. J Vet Diagn Invest 8:124–130

Tamburrini E, Ortona E, Visconti E et al (1999) *Pneumocystis carinii* infection in young non-immunosuppressed rabbits. Kinetics of infection and of the primary specific immune response. Med Microbiol Immunol 188:1–7

Tang X, Bartlett MS, Smith JW et al (1998) Determination of copy number of rRNA genes in *Pneumocystis carinii* f. sp. *hominis*. J Clin Microbiol 36:2491–2494

Taylor JW, Jacobson DJ, Kroken S et al (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31:21–32

Thomas CF Jr, Limper AH (2007) Current insights into the biology and pathogenesis of *Pneumocystis* pneumonia. Nat Rev Microbiol 5:298–308

Thomas CF, Anders RA, Gustafson MP et al (1998) *Pneumocystis carinii* contains a functional cell-division-cycle Cdc2 homologue. Am J Respir Cell Mol Biol 18:297–306

Totet A, Latouche S, Lacube P et al (2004) *Pneumocystis jirovecii* dihydropteroate synthase genotypes in immunocompetent infants and immunosuppressed adults, Amiens, France. Emerg Infect Dis 10:667–673

Utili R, Durante-Mangoni E, Basilico C et al (2007) Efficacy of caspofungin addition to trimethoprim-sulfamethoxazole treatment for severe *Pneumocystis* pneumonia in solid organ transplant recipients. Transplantation 84:685–688

Vargas SL, Ponce CA, Gigliotti F et al (2000) Transmission of *Pneumocystis carinii* DNA from a patient with *P. carinii* pneumonia to immunocompetent contact health care workers. J Clin Microbiol 38:1536–1538

Vasquez J, Smulian AG, Linke MJ et al (1996) Antigenic differences associated with genetically distinct *Pneumocystis carinii* from rats. Infect Immun 64:290–297

Vavra J, Kucera K (1970) *Pneumocystis carinii* D?elanoë, its ultrastructure and ultrastructural affinities. J Protozool 17:463–483

Villegas LR, Kottom TJ, Limper AH (2009) Characterization of PCEng2 a {beta}-1,3-endoglucanase homologue in *Pneumocystis carinii* with activity in cell wall regulation. Am J Respir Cell Mol Biol 43(2):192–200

Vohra PK, Puri V, Thomas CF Jr (2003a) Complementation and characterization of the *Pneumocystis carinii* MAPK, PCM. FEBS Lett 551:139–146

Vohra PK, Puri V, Kottom TJ et al (2003b) *Pneumocystis carinii* STE11, an HMG-box protein, is phosphorylated by the mitogen activated protein kinase PCM. Gene 312:173–179

Vohra PK, Park JG, Sanyal B et al (2004) Expression analysis of PCSTE3, a putative pheromone receptor from the lung pathogenic fungus *Pneumocystis carinii*. Biochem Biophys Res Commun 319:193–199

Wakefield AE (1998) Genetic heterogeneity in *Pneumocystis carinii*: an introduction. FEMS Immunol Med Microbiol 22:97–101

Wakefield AE (2002) Pneumocystis carinii. Br Med Bull 61:175–188

Wakefield AE, Peters SE, Banerji S et al (1992) *Pneumocystis carinii* shows DNA homology with the ustomycetous red yeast fungi. Mol Microbiol 6:1903–1911

Wakefield AE, Keely SP, Stringer JR et al (1997) Identification of porcine *Pneumocystis carinii* as a genetically distinct organism by DNA amplification. APMIS 105:317–321

Weinberg GA, Bartlett MS (1991) Comparison of pulsed field gel electrophoresis karyotypes of *Pneumocystis carinii* derived from rat lung, cell culture and ferret lung. J Protozool 38:64–65

Wilson DE, Bartsch RC, Bigalke RD et al (1974) Observations on mortality rates and disease in roan and sable antelope on nature reserves in the Transvaal. J S Afr Wildl Manage Assoc 4:203–206

Wyder MA, Rasch EM, Kaneshiro ES (1998) Quantitation of absolute Pneumocystis carinii nuclear DNA content. Trophic and cystic forms isolated from infected rat lungs are haploid organisms. J Eukaryot Microbiol 45:233–239

Yonushonis WP, Elwell MR, Lawyer PG et al (1986) Infection of a three-toed sloth (*Bradypus variegatus*) by a *Pneumocystis*-like organism in Panama. Wildl Dis 22:572–575

Yoshida Y (1989) Ultrastructural studies of *Pneumocystis carinii*. J Protozool 36:53–60

Yoshida Y, Yamada M, Shiota T et al (1981) Provocation experiment: *Pneumocystis carinii* in several kinds of animals. Zentralbl Bakteriol Mikrobiol Hyg A 250:206–212

Yoshida Y, Matsumoto Y, Yamada M et al (1984) *Pneumocystis carinii*: electron microscopic investigation on the interaction of trophozoite and alveolar lining cell. Zentralbl Bakteriol Mikrobiol Hyg A 256:390–399

Chapter 9
Hantavirus Genetic Diversity

Kim Blasdell, Heikki Hentonnen, and Philippe Buchy

Abstract The genus *Hantavirus*, belonging to the family *Bunyaviridae*, is a rapidly expanding group of viruses. Although some members of the genus have been well known for over 30 years, in the last 5 years at least 12 putative novel hantavirus species have been identified, primarily in soricomorph mammals. Although just over 20 species of mainly rodent-borne hantaviruses are officially recognized, these newly identified species are likely increase the real total to between 30 and 50 species, depending on the parameters used to define a species. Many hantaviruses are important zoonotic pathogens, responsible for up to 200,000 human cases annually. Although the earliest hantaviruses to be recognized were usually identified in humans, and rodent samples were often differentiated through serological techniques, in the past 20 years almost all of the hantavirus species were identified using molecular methods. The molecular tools now available have allowed a more precise characterisation of hantavirus taxonomy, hantavirus species distribution,

K. Blasdell (✉)
Virology Unit, Institut Pasteur du Cambodge, 5 Monivong Boulevard,
Phnom Penh BP983, Cambodia

Institut des Sciences de l'Evolution, CNRS, IRD, Université Montpellier 2,
CC065, 34095 Montpellier Cedex 05, France

Australian Animal Health Laboratory, CSIRO, 5 Portarlington Road, Geelong,
VIC 3219, Australia
e-mail: kim.blasdell@csiro.au

H. Hentonnen
Vantaa Research Unit, Finnish Forest Research Institute, POB 18,
Vantaa FI-01301, Finland
e-mail: heikki.henttonen@metla.fi

P. Buchy
Virology Unit, Institut Pasteur du Cambodge, 5 Monivong Boulevard,
Phnom Penh BP983, Cambodia
e-mail: pbuchy@pasteur-kh.org

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_9,

the identification of highly genetically different hantavirus species, recombination within and between hantavirus species and the genetic predisposition to severe hantavirus disease in humans.

Abbreviations

aa	Amino acid
BAYV	Bayou virus
CFR	Case fatality rate
DOBV	Dobrava virus
HCPS	Hantavirus Cardio-Pulmonary Syndrome
HLA	Human leukocyte antigen
HNTV	Hantaan Virus
HFRS	Hemorrhagic Fever with Renal Syndrome
ICTV	International Committee on Taxonomy of Viruses
L	Large (segment)
M	Medium (segment)
MHC	Major histocompatibility complex
nt	nucleotide
PUUV	Puumala virus
RT-PCR	Reverse transcription polymerase chain reaction
S	Small (segment)
SEOV	Seoul virus
SNV	Sin Nombre virus
THAIV	Thailand virus
TULV	Tula virus

9.1 Introduction

Hantaviruses are single-stranded, negative-sense, tri-segmented RNA viruses in the genus *Hantavirus*, family *Bunyaviridae* (Schmaljohn and Hooper 2001). Present in Europe, Asia, Africa and the Americas, some hantaviruses cause human disease, with thousands of cases recorded each year (Song et al. 1984). In several locations in Europe, the Americas and Asia, hantavirus disease is endemic, with both seasonal and often multi-annual fluctuations associated with host population variations, often related to environmental and possibly to climatic variations (Clement et al. 2009; Ferrer et al. 2003; Hjelle and Glass 2000; Zou et al. 2008a). Several new foci of hantaviruses have been identified in recent years, many of which have been linked to shifting land-use patterns and human encroachment (Mills 2006).

All hantaviruses are primarily hosted by a single rodent or soricomorph species (Kang et al. 2009c; Plyusnin and Morzunov 2001). While most viruses of the *Bunyaviridae* family require arthropods for transmission, hantaviruses do not, although some surveys have identified hantaviruses in arthropods (Deng et al. 2008;

Houck et al. 2001). Transmission between rodents occurs through aerosolised excreta (Kallio et al. 2006; Houck et al. 2001) and through aggressive encounters (Escutenaire et al. 2002; Glass et al. 1988; Mills et al. 1999). Hantaviruses are hosted by several rodent sub-families within the families Cricetidae and Muridae. Mice and rats in the subfamily Murinae host hantaviruses almost worldwide, while hantaviruses are hosted by Arvicolinae rodents in North America and Eurasia and by Sigmodontinae and Neotominae rodents in North and South America (Herbreteau et al. 2006). Soricid and Talpid-borne hantaviruses have so far been detected in Europe, Asia, Africa and North America (Arai et al. 2007, 2008a, 2008b; Kang et al. 2009b, c; Song et al. 2007a, b, c, 2009).

Although Thottapalayam virus, a shrew-borne virus, was the first hantavirus to be detected (Carey et al. 1971), until recently all other hantaviruses had been detected in rodents. In many cases the phylogeny of the viruses appears to mirror that of their rodent hosts, with a strong concordance at the genus and subfamily level (Herbreteau et al. 2006), suggesting that the rodents and their viruses may have co-evolved. The recent discoveries of several other soricomorph-borne hantaviruses, and their resulting phylogenetic positions, have called this hypothesis into question (Ramsden et al. 2009). A more recent hypothesis has tried to explain the current findings by a combination of host-switching and local adaptation, alongside some co-evolution (Kang et al. 2009c). However, these new questions are far from being answered with certainty and virtually nothing is known about possible soricomorph-borne hantavirus infections in humans.

9.2 Overview

9.2.1 Virus Morphology and Genome Organisation

Hantavirus virions are spherical and 80–120 nm in diameter. Surface glycoproteins, projecting 5–10 nm, are embedded in a 5–7 nm thick lipid bi-layered envelope (Schmaljohn and Hooper 2001). In contrast to other members of the *Bunyaviridae*, the surface structure of *Hantavirus* virions display a square grid-like appearance (McCormick et al. 1982). The virion interior contains the three ribonucleocapsids and the L (polymerase) protein and therefore shows a filamentous or coiled bead-like appearance (Hung et al. 1985) (see Fig. 9.1).

The three segments are designated small (S), medium (M) and large (L) and are approximately 1,500–2,000 bp, 3,600 bp and 6,000–6,500 bp in length, respectively. The S RNA encodes the nucleocapsid protein, the L RNA the L protein (an RNA-dependent RNA polymerase) and the M RNA encodes a polyprotein that is cleaved during translation to produce the envelope glycoproteins, G1 and G2 (Schmaljohn and Hooper 2001). A hypothetical non-structural ORF has also been identified in the S segment of some hantavirus species, primarily in Arvicoline, Sigmodontini and Neotomini-borne hantavirus species. This protein is found in a reading frame that overlaps with the nucleocapsid protein reading frame (Jaaskelainen

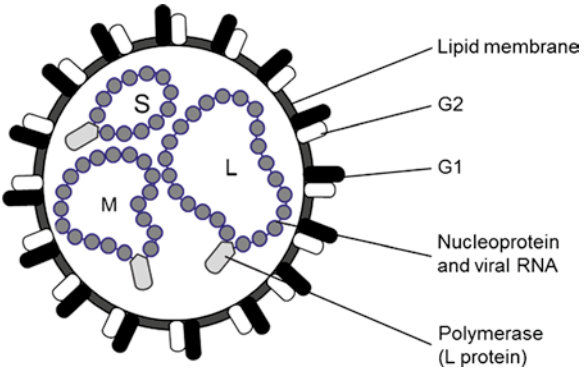

Fig. 9.1 Morphology of *Hantavirus* virions

et al. 2007). All three RNA segments have the same complementary nucleotides at their 3′ and 5′ ends and this sequence is conserved for all viruses within the *Hantavirus* genus. This is also true for other members of the *Bunyaviridae*, although the terminal sequences differ between the genera (Schmaljohn and Hooper 2001).

9.2.2 Human Disease

In humans, most hantavirus infections manifest in one of two distinct syndromes: Hemorrhagic Fever with Renal Syndrome (HFRS), often referred to as nephropathia epidemica (NE) in the milder Puumala virus induced HFRS in Europe, and Hantavirus Cardio-Pulmonary Syndrome (HCPS) observed in the Americas. Transmission to humans normally occurs through inhalation of infected aerosolised rodent excreta (Pettersson et al. 2008).

Hemorrhagic Fever with Renal Syndrome is present throughout much of Eurasia and has also been reported in Africa (Bi et al. 2008; Coulaud et al. 1987) and sporadically elsewhere (Glass et al. 1994; Seijo et al. 2003). Several hantaviruses cause this disease, including Puumala virus (PUUV), Seoul virus (SEOV), Hantaan virus (HTNV), Saaremaa virus (SAAV), Dobrava virus (DOBV) (Heyman and Vaheri 2008; Kim et al. 1995). After contamination, the incubation period in humans usually lasts from 1 to 3 weeks (Kramski et al. 2009). This is followed by the onset of variable, often influenza-like symptoms. The duration of this phase varies typically between 3 and 7 days. A hypotensive phase lasting several hours to days then occurs, which may also result in shock and renal failure contributing to half of the deaths caused by this disease. The marked thrombocytopenia usually observed participates in haemorrhage, hence the classification as a form of hemorrhagic fever. Increased capillary permeability is likely responsible for the retroperitoneal oedema and the lower back pain often described (Vinh and Embil 2009). Most of the human cases in Europe are caused by PUUV, which usually results in a mild form of HFRS (nephropathia epidemica) with a low case fatality rate (CFR) of less than 1%. DOBV, which has a more restricted geographic range in south east Europe, is associated

with more severe disease (up to 15% CFR) (Heyman and Vaheri 2008). SAAV has been implicated in mild PUUV-like human disease, and so also has Tula virus (TULV) although the link is questionable for the latter (Golovljova et al. 2007; Schultze et al. 2002). In Asia, most human cases are caused either by HTNV, where disease often follows a severe course (up to 5% CFR) or SEOV, where disease is usually mild to moderate (1–2% mortality) (Schmaljohn and Hooper 2001; Vapalahti et al. 2003).

Hantavirus Cardio-Pulmonary Syndrome (HCPS) is restricted to the Americas and was discovered less than 20 years ago, although disease has probably been long known in native tradition (Jalbert 1997). Sin Nombre virus (SNV) and Bayou virus (BAYV) in North America (Morzunov et al. 1995; Rivers et al. 2009) and various lineages of Andes virus (ANDV) in South and Central America (Lopez et al. 1996; Vincent et al. 2000) have been associated with HCPS. Case fatality rates are high, sometimes reaching 50% during some small outbreaks (Bi et al. 2008; Rivers et al. 2009). The incubation period ranges from 9 to 33 days. During the initial phase, malaise and fever are observed, but as the illness progresses, the patients develop cough and worsening dyspnea. Interstitial pulmonary oedema resulting from the increased capillary permeability may be detected radiographically and this may lead to respiratory failure (Vinh and Embil 2009).

9.2.3 Laboratory Approach

Hantaviruses are genetically very diverse and this can be problematic for the identification of new species, particularly for species non-pathogenic in humans. Originally hantaviruses were differentiated using serological techniques such as cross-neutralisation techniques, often in combination with culture in cells and/or mice. However, a great deal of serological cross-reaction occurs between species and particularly between those hosted by rodents of the same family (Asada et al. 1989). Since the 1990s, molecular techniques, and especially reverse-transcriptase polymerase chain reaction (RT-PCR), have been regularly employed in the identification of hantaviruses (Table 9.1). Other tools such as standard multiplex RT-PCRs or real-time RT-PCR which allows virus identification and viral load quantification (Korva et al. 2009b) have also been developed (Kramski et al. 2007). Random priming and pyrosequencing have been successfully used for the genetic characterisation of some of the more diverse members of the *Hantavirus* genus (Kramski et al. 2007; Lokugamage et al. 2002; Song et al. 2009).

9.3 Virus Species

The use of molecular techniques and subsequent genetic analysis has undoubtedly improved the ability to identify novel hantavirus species and allowed quicker and more accurate classification of novel viruses and subscription to particular species groups.

Table 9.1 Hantavirus species, primary host, geographic range, relationship to other viruses and human disease, genetic diversity within species and ICTV status

Species	Host species	Geographic distribution	Potential variants	Human disease	Diversity within species at gene segment level	Year and method of identification	ICTV confirmed
Hantaan virus	*Apodemus agrarius*	Northern Asia	Unknown virus, India, (Chandy et al. 2009a)	HFRS	25% nt, 16% aa, entire M segment, (Wang et al. 2000)	1978, isolation in reservoir host, (Lee et al. 1978)	Yes
Seoul virus	*Rattus norvegicus*	Worldwide	N/A	HFRS	16% nt, 3.9% aa, entire M segment, (Wang et al. 2000)	1982, isolation in Wistar rats, (Lee et al. 1982b)	Yes
Dobrava virus	*Apodemus* spp.	Central and eastern Europe, Russia	May form single species (Dobrava-Belgrade virus), (Maes et al. 2009)	HFRS	21.7% nt, 9.8% aa, entire M segment, (Klempa et al. 2008)	1992, molecular methods and isolation from reservoir, (Avsic-Zupanc et al. 1992)	Both confirmed
Saaremaa virus	*Apodemus agrarius*	Central and eastern Europe		Mild HFRS		1999, molecular methods and isolation in cell culture from reservoir, (Nemirov et al. 1999)	
Thailand virus	*Bandicota indica*	Thailand	Possibly present in India, (Chandy et al. 2008)	Possible HFRS	3.5% nt, entire S segment, (Hugot et al. 2006)	1992, cell culture and molecular methods from reservoir, (Arthur et al. 1992)	Under review
Sangassou virus	*Hylomyscus simus*	Guinea	N/A	Unknown	Only one strain available, (Klempa et al. 2006)	2006, molecular methods from reservoir, (Klempa et al. 2006)	No

Serang virus	*Rattus tanezumi*	Indonesia	Partially sequenced viruses from Cambodia and Singapore, (Johansson et al. 2010; Reynes et al. 2003)	Unknown	Only one strain available, (Plyusnina et al. 2009b)	2006, molecular methods from reservoir, (Plyusnina et al. 2009b)	No
Puumala virus	*Myodes glareolus*	Eurasia	Muju virus (variant from Korea), Hokkaido virus (in *M. rufocanus* from Japan), (Daud et al. 2007)	Mild HFRS	23.7% nt entire S segment, 11.6% entire M segment, (Song et al. 2007a)	1980, detection of antigen in reservoir and of antibodies in human patients, (Brummer-Korvenkontio et al. 1980)	Yes
Prospect Hill virus	*Microtus pennsylvanicus*	USA	N/A	None	7.7% nt entire S segment (Genbank)	1982, isolation from reservoir, (Lee et al. 1982a)	Yes
Tula virus	*Microtus arvalis* and other sympatric *Microtus* species in western palearctic	Europe and central Asia	N/A	Potentially HFRS	16% nt entire S segment, (Plyusnina et al. 2008b)	1994, molecular methods from reservoir, (Plyusnin et al. 1994)	Yes
Isla Vista virus	*Microtus californicus*	California, USA	N/A	None	11% nt partial M segment (Song et al. 1995)	1995, molecular methods from reservoir, (Song et al. 1995)	Yes

(continued)

Table 9.1 (continued)

Species	Host species	Geographic distribution	Potential variants	Human disease	Diversity within species at gene segment level	Year and method of identification	ICTV confirmed
Khabarovsk virus	*Microtis fortis*	Russia and China	May form a single species,	None	4.4% aa entire M segment, (Zou et al. 2008c)	1996, molecular methods from reservoir hosts, (Horling et al. 1996)	Yes
Topografov virus	*Lemmus sibiricus*	Russia and China		None	7.5% nt partial S segment, (Vapalahti et al. 1999b)	1996, detection of antigen in reservoir and isolation, (Plyusnin et al. 1996b)	Yes
Vladivostok virus	*Microtis fortis*	Russia and China	May be related to the above	None	18% nt and 5% aa complete S segment, (Zou et al. 2008d)	1999, molecular methods from reservoir, (Kariwa et al. 1999)	No
Sin Nombre virus	*Peromyscus maniculatus*	North America	New York and Monogahela variants, also Blue River virus,	HCPS	20.1% nt and 6.6% aa, entire M segment, (Morzunov et al. 1998)	1994, isolation in mice and cell culture, molecular methods from reservoir, (Elliot et al. 1994)	Yes
Rio Segundo virus	*Reithrodontomys* spp. and *Peromyscus boylii*	Americas	El Moro Canyon and Limestone Canyon variants, (Hjelle et al. 1994; Sanchez et al. 2001)	None	27.3% nt and 19.7% aa partial M segment, (Sanchez et al. 2001)	1994, molecular methods from reservoir, (Hjelle et al. 1995a)	Yes

Bayou virus	*Oryzomys palustris*	Southern USA and Mexico	Black Creek Canal and Muleshoe viruses hosted by *Sigmodon hispidus*, Catacamas virus and Playa de Oro viruses hosted by *Oryzomys couesi*, (Chu et al. 2008; Milazzo et al. 2006)	HCPS	24.4% nt entire S segment, 12.5% GP1 protein, (Ravkov et al. 1995)	1995, molecular methods from fatal human case, (Morzunov et al. 1995)	Yes
Andes virus	*Oligoryzomys* spp. and others	South America	Many lineages	HCPS	24.5% nt and 8.1%, partial M segment, (Padula et al. 2000, 2002b)	1996, molecular methods from fatal human case, (Lopez et al. 1996)	Yes
Cano Delgadito virus	*Sigmodon alstoni*	Venezuela	N/A	Unknown	Only one strain available	1997, molecular methods and isolation from reservoir, (Fulhorst et al. 1997)	Yes
Thottapalayam virus	*Suncus murinus*	Southern Asia	N/A	Possible disease	Only one strain available	1971, isolated in cell culture from soricid hosts, (Carey et al. 1971)	Yes
Tanganya virus	*Crocidura theresea*	Guinea	N/A	Unknown	Only one strain available	2007, molecular methods from reservoir, (Klempa et al. 2007)	No

(continued)

Table 9.1 (continued)

Species	Host species	Geographic distribution	Potential variants	Human disease	Diversity within species at gene segment level	Year and method of identification	ICTV confirmed
Camp Ripley virus	*Blarina brevicauda*	USA	N/A	Unknown	1.9% nt, partial M segment, (Arai et al. 2007)	2007, molecular methods from reservoir, (Arai et al. 2007)	No
Seewis virus	*Sorex araneus*	Europe, Siberia	N/A	Unknown	21.8% nt, partial L segment, (Kang et al. 2009a)	2007, molecular methods from reservoir, (Song et al. 2007b)	No
Cao Bang virus	*Anourosorex squamipes*	Vietnam	N/A	Unknown	Not specified	2007, molecular methods from reservoir, (Song et al. 2007c)	No
Ash River virus	*Sorex cinereus*	USA	N/A	Unknown	Only one strain available	2008, molecular methods from reservoir, (Arai et al. 2008a)	No
Jemez Springs virus	*Sorex monticolus* and other species	USA	N/A	Unknown	12.1% nt and 0.9% aa, partial L segment, (Arai et al. 2008a)	2008, molecular methods from reservoir, (Arai et al. 2008a)	No
Asama virus	*Urotriches talpoides*	Japan	N/A	Unknown	1.1% nt and 0.7% aa, entire M segment, (Arai et al. 2008b)	2008, molecular methods from reservoir, (Arai et al. 2008b)	No

Oxbow virus	*Neurotrichus gibbsii*	USA	N/A	Unknown	Only one strain available	2009, molecular methods from reservoir, (Kang et al. 2009b)	No
Imjin virus	*Crocidura lasiura*	Korea	N/A	Unknown	9.3% nt and 0.7% aa, entire S segment, (Song et al. 2009)	2009, molecular methods and isolation from reservoir, (Song et al. 2009)	No
Nova virus	*Talpa europaea*	Europe	N/A	Unknown	Only one strain available	2009, molecular methods from reservoir, (Kang et al. 2009c)	No

HFRS Hemorrhagic Fever with Renal Syndrome; *HCPS* Hantavirus Cardio-Pulmonary Syndrome; *nt* nucleotide; *aa* amino acid

However, the definition of what actually constitutes a separate species in terms of nucleic acid and amino acid diversity is still under much debate. Several criteria, defined by the International Committee on Taxonomy of Viruses (ICTV) need to be met in order for novel isolates to be classified as a separate species (Nichol et al. 2005). At present, the guidelines for hantaviruses are comprised of four rules:

1. A hantavirus species is found in a unique ecological niche, i.e. a different primary host species or subspecies.
2. A minimum of 7% difference at the amino acid level is displayed between the novel isolate and previously identified viruses, when comparing both entire M and S segments.
3. Separate hantavirus species shows a minimum of a fourfold difference in two-way cross-neutralisation tests.
4. Re-assortments between hantavirus species do not naturally occur.

However several hantaviruses that have been confirmed as separate species by the ICTV do not fulfil the second rule and with the identification of more and more putative novel hantavirus species, it is clear that this rule at least is in need of review. A recent paper has suggested that the difference in amino acid sequence between species should be raised to a minimum of 10% and 12% for the entire S and M segments, respectively (Maes et al. 2009). If this proposition is accepted, it would considerably reduce the number of hantavirus species, several of which have been clearly demonstrated to have different reservoir host species. This is complicated further by the definition of what a "host species" is. Although several different species may become naturally infected, many of them will be secondary or spill-over hosts, which may or may not contribute to the epidemiology of the virus. In this chapter the primary reservoir host of each hantavirus species will be stated and although putative secondary hosts are also mentioned, their role in virus transmission will not be defined.

Currently over 20 species of hantavirus are officially recognised and most of these are sub-divided into several sub-species and lineages, although the taxonomy is constantly undergoing revision. Phylogenetic methods infer that there are at least five sub-groups, including one group hosted by murine rodents, one by arvicoline rodents, a third by sigmodontine and neotomine rodents and the remaining two groups by small mammals in the order Soricomorpha (Kang et al. 2009c; Maes et al. 2009) (see Fig. 9.2).

9.3.1 Murine-Borne Viruses

Hantaan virus was the first hantavirus pathogenic for humans to be recognised. Although the human disease was known for a long time, it was not until 1978 that this virus was successfully isolated from its primary reservoir host, *Apodemus agrarius* (Lee et al. 1978). Alongside SEOV, HNTV is responsible for the majority of the hantavirus-associated human disease in Asia (Song 1999). HNTV has been detected

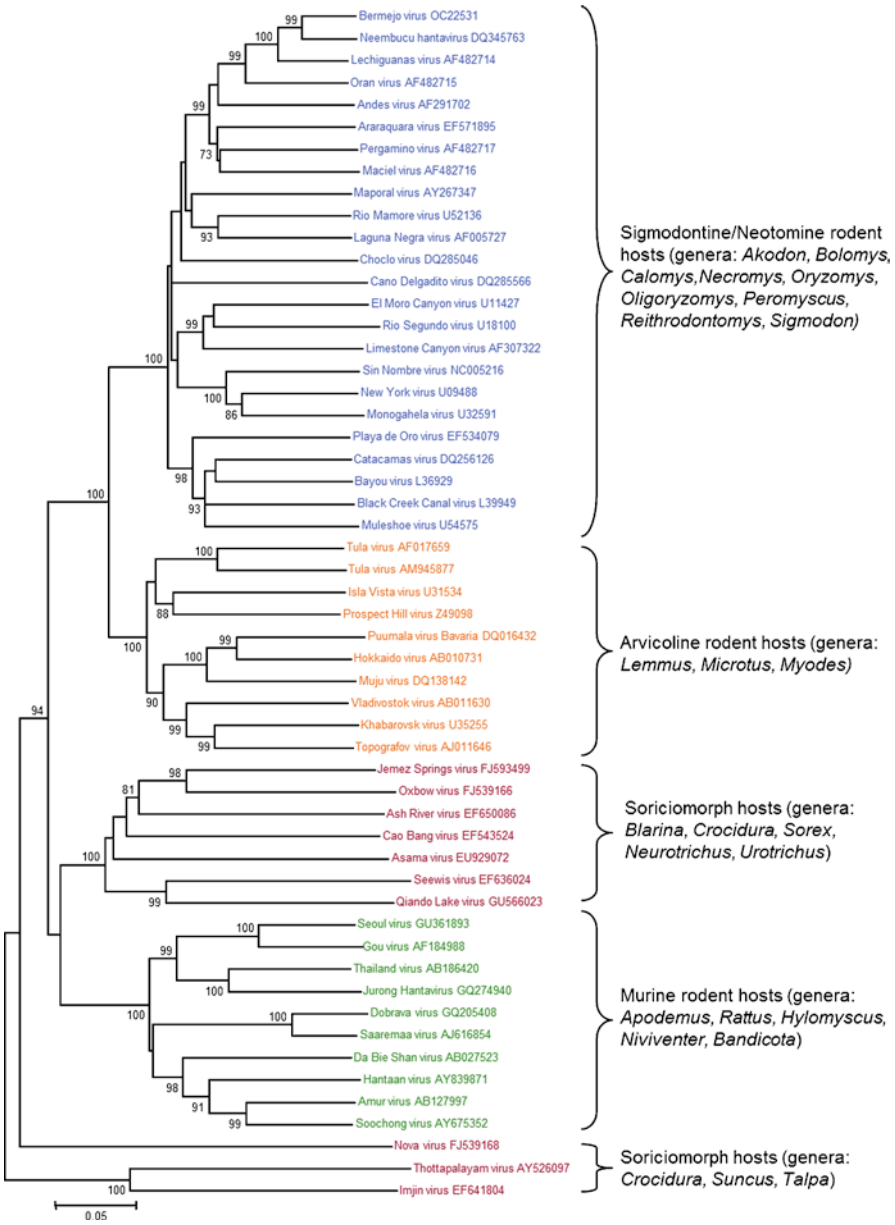

Fig. 9.2 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial S segment, showing bootstrap values ≥70 for representatives of most hantavirus species. GenBank accession numbers are shown next to each sequence. Proposed reservoir host sub-families are also noted. Branch lengths are to scale

throughout much of China, Korea and parts of north-eastern Russia (Lee 1989; Lokugamage et al. 2004; Song 1999). It has been found primarily in rural areas due to the habitat preference of its associated rodent host species (Chan et al. 1987) and is divided into at least three lineages (A, B and C), each carried by a different rodent host species, namely *Apodemus agrarius, A. peninsulae* and *Niviventer confucianus* respectively. Lineage A is found throughout Korea and China. Lineage B, including Amur and Soochong viruses, is found in northeastern China and Russia. Lineage C , also known as Da Bie Shan virus, was identified in *Niviventer confucianus* from Anhui province in eastern China (Maes et al. 2009) (Fig. 9.3).

Seoul virus is probably the most geographically widespread hantavirus. Discovered in 1982, this hantavirus causes HFRS in humans (Lee et al. 1982b). Believed to have an Asian origin, SEOV has been detected in many countries and cities throughout the world due to its association with its host, *Rattus norvegicus* (Cueto et al. 2008; Easterbrook et al. 2007; Glass et al. 1994; Heyman et al. 2004, 2009), known as the brown or Norway rat, which itself is distributed almost globally. It has also been associated with *Rattus rattus* and *Mus musculus* in some cities in China (Wang et al. 2000; Zhang et al. 2009) and is responsible for thousands of human disease cases annually (Chen et al. 1986; Song 1999).

Dobrava and Saaremaa viruses have been identified more recently (in 1992 and 1999, respectively) (Avisc-Zupanc et al. 1992; Nemirov et al. 1999) and some authors suggest that these two viruses actually form a single species, known as Dobrava-Belgrade virus (Maes et al. 2009; Nemirov et al. 2002; Plyusnin et al. 2003). There are several differences between the two viruses on both a serological (Sjolander et al. 2002) and molecular level (Sironen et al. 2005). They are also carried by different rodent hosts and while infection with DOBV causes severe human disease (with CFR of up to 15%) (Heyman and Vaheri 2008), human infection with SAAV results in mild or asymptomatic disease (Golovljova et al. 2002; Plyusnin et al. 2006) and has not been associated with any human fatalities. Different lineages of DOBV – SAAV complex are associated with different species of *Apodemus* mice: DOBV in *A. flavicollis* in south eastern Europe, two lineages of SAAV in *A.agrarius* in central and eastern Europe and DOBV-Shotsi in *A. ponticus* in south western Russia (Avsic-Zupanc et al. 1992; Klempa et al. 2008; Sibold et al. 2001; Tkachenko et al. 2005). These different lineages also appear to be associated with different degrees of disease severity. SAAV was identified in *A. agrarius* in north eastern Europe and western Russia (Nemirov et al. 2002) although occasional spill-over infections have also been detected in *A. flavicollis* (Schlegel et al. 2009).

Thailand virus (THAIV) was described in the early 1990s (Xiao et al. 1994). This species is associated with the greater Bandicoot rat, *Bandicota indica*. So far the virus has only been definitively identified in Thailand (Hugot et al. 2006), although its host species is wide-spread throughout south and southeast Asia (Aplin et al. 2003) and a recent report from India suggests that it may also be present there (Chandy et al. 2009a). So far, only a single non-fatal human case from Thailand and two serologically-identified human cases from India have been linked to THAIV infection (Chandy et al. 2009a; Pattamadilok et al. 2006).

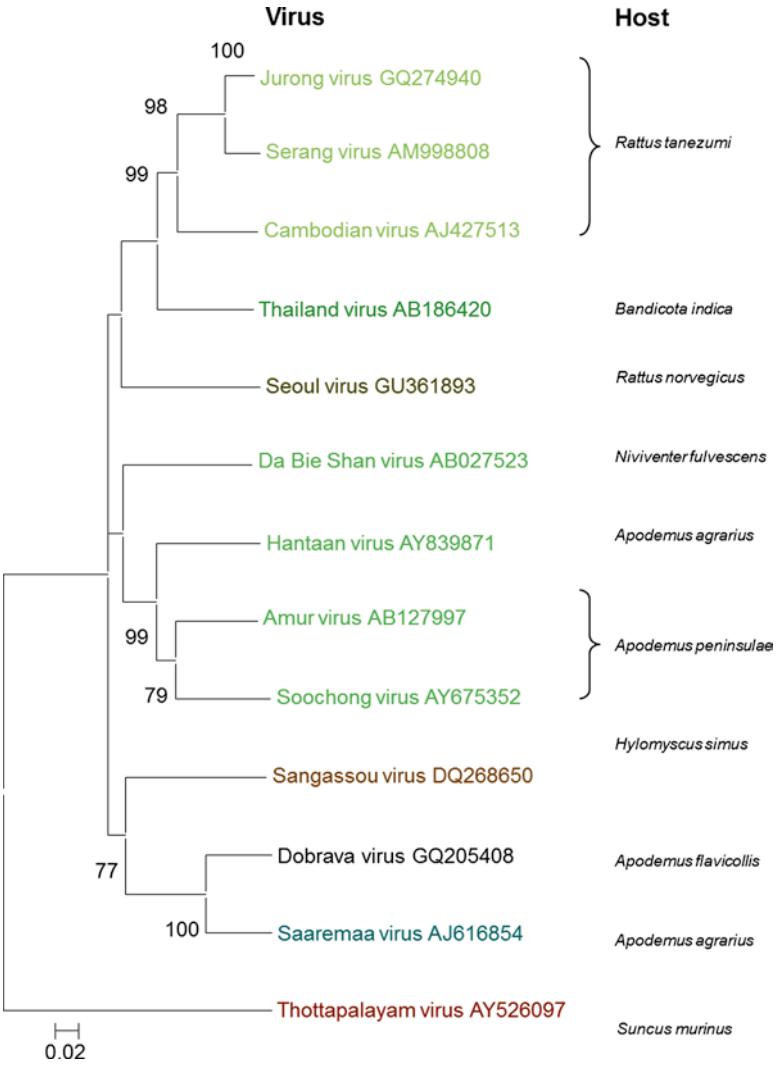

Fig. 9.3 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial S segment, showing bootstrap values ≥70 for representatives of all murine-borne hantavirus species and rooted to Thottapalayam virus. GenBank accession numbers and proposed primary rodent hosts are shown next to each sequence. Branch lengths are to scale

Sangassou virus was identified in 2006 in an African wood mouse, *Hylomyscus simus*, in Guinea (Klempa et al. 2006). There have been reports of HFRS-like disease in sub-Saharan Africa (Coulaud et al. 1987) including in the village where Sangassou virus was originally isolated (Klempa et al. 2010) but the virus has never been isolated from a patient.

Serang virus is the most recent murine-borne hantavirus to be identified and was discovered in Indonesia in 2009. This virus species appears to be hosted by the Asian house rat, *Rattus tanezumi*, a member of the *Rattus rattus* species complex (Plyusnina et al. 2009b). So far this virus has only been identified in rodents and it may be related to a partially characterised virus detected previously in *Rattus tanezumi* in Cambodia (Reynes et al. 2003) and to another partially characterised virus recently found in Singapore (Johansson et al. 2010).

9.3.2 Arvicoline-Borne Viruses

Puumala virus, the first arvicoline-borne hantavirus found to cause disease in humans, was identified in its rodent hosts in Finland in 1980 (Brummer-Korvenkontio et al. 1980). Thousands of human cases in Europe are reported annually. The incidence varies both seasonally and multi-annually, associated with fluctuations in the abundance of the reservoir host, which is in turn related to predator driven rodent cycles in boreal zones and masting events in temperate zones (Dixon 2009; Linard et al. 2007). Some studies have suggested that climate change may alter the incidence of this disease (Schwarz et al. 2009; Tersago et al. 2009). PUUV is carried by the arvicoline rodent, *Myodes glareolus*, the bank vole, throughout much of Europe, Russia and into western Siberia, and by *M. rufocanus*, the grey- sided vole, in easternmost Siberia, Korea and Japan (Brummer-Korvenkontio et al. 1980; Chumakov et al. 1981; Sommer et al. 1985). It is divided into three major lineages, that containing the European form of PUUV found throughout the whole range of *M. glareolus*, the second lineage containing Muju virus, the Puumala variant identified in *Myodes regulus* in Korea (Maes et al. 2009; Song et al. 2007d) and the third consisting of Hokkaido virus, identified in *Myodes rufocanus* in Japan, and the Far East to Lake Baikal in Central Siberia (Daud et al. 2007; Plyusnina et al. 2008a) (Fig. 9.4).

Prospect Hill virus was the first hantavirus to be identified in the New World in 1982 (Lee et al. 1982a). It has been found in meadow voles, *Microtus pennsylvanicus*, in several states throughout the United States (Lee et al. 1985; Yanagihara et al. 1987). Although sero-studies carried out before the discovery of SNV and HCPS suggested that Prospect Hill virus was the causative agent of sero-conversions in humans (Forthal et al. 1987), this virus has never been connected to human disease resembling other *Microtus*-borne viruses such as TULV in Eurasia.

Tula virus is the only other arvicoline-borne hantavirus that has been implicated as a potential human pathogen. This species was identified relatively recently, in the early 1990s, in European common voles, *Microtus arvalis* and *M. levis* (formerly *rossiaemeridionalis*) from the Tula region in Russia (Plyusnin et al. 1994). It has since been found in voles as far east as eastern Kazakhstan (Plyusnina et al. 2008b) and as far west as the Netherlands (Reusken et al. 2008) and Belgium (Heyman et al. 2002), with Kazakh strains being the most divergent (Plyusnina et al. 2008b). Its primary host is *Microtus arvalis* but TULV can also infect several other sympatric

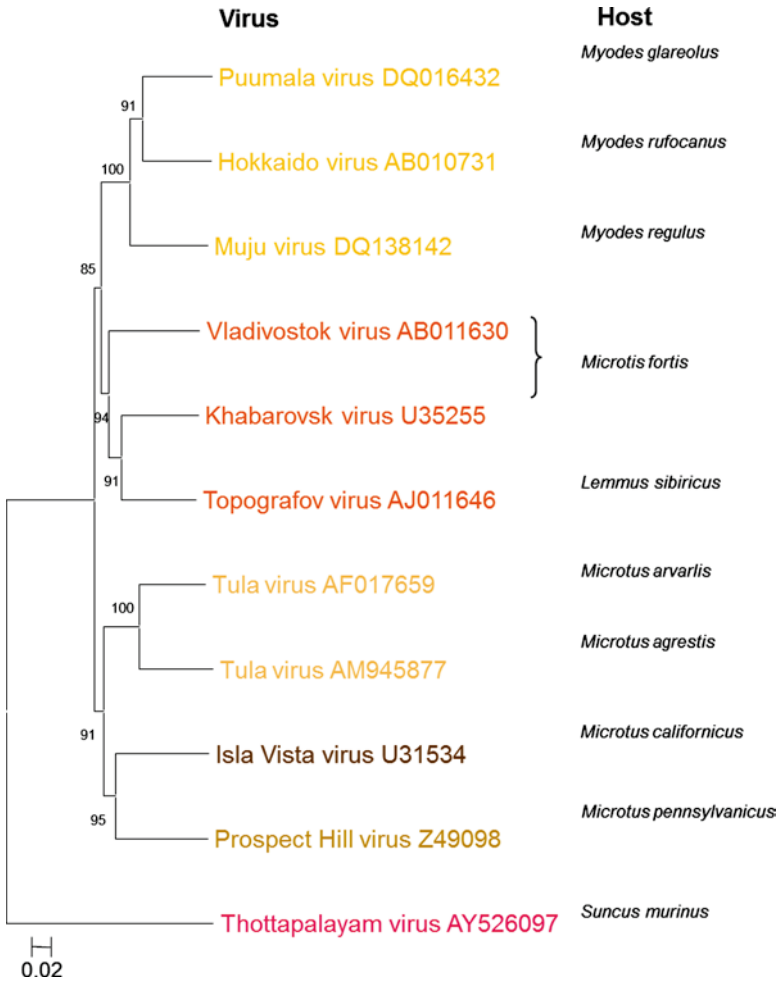

Fig. 9.4 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial S segment, showing bootstrap values ≥70 for representatives of all arvicoline-borne hantavirus species and rooted to Thottapalayam virus. GenBank accession numbers and proposed primary rodent hosts are shown next to each sequence. Branch lengths are to scale

Microtus species, including *Microtus agrestis*, as seen in Croatia and Germany (Plyusnina et al. 2008b; Scharninghausen et al. 2002; Schmidt-Chanasit et al. 2010). Although some hantavirus infections confirmed by serology have been observed in a region where Tula virus circulates in rodents (Klempa et al. 2003a; Schultze et al. 2002), its pathogenicity for humans has not yet been proven.

Isla Vista virus is another New World arvicoline-borne hantavirus, identified in 1995 in *Microtus californicus*, the Californian vole (Song et al. 1995). So far this virus species has only been found in the state of California and has not been implicated in any human disease.

Khabarovsk virus and Topografov virus, isolated from *Microtis fortis* in 1996 (Horling et al. 1996) and *Lemmus sibiricus* in 1999 (Plyusnin et al. 1996b; Vapalahti et al. 1999a) respectively, have both been identified from Russia. A variant of Khabarovsk was also discovered in *Microtus maximowiczii* in China (Zou et al. 2008c). A recent revision of hantavirus taxonomy has implied that they are two lineages of the same virus species (Maes et al. 2009). A host switch by Topografov virus to *Lemmus sibiricus* has been proposed, as both Khabarovsk virus and the closely related Vladivostok virus (found in eastern Asia (Plyusnina et al. 2008a; Zou et al. 2008d)), with which Topografov virus forms a monophyletic group (Fig. 9.4), are carried by *Microtus* voles (Vapalahti et al. 1999a, b). However, this monophyletic group does not group with other *Microtus*-borne hantaviruses but forms a sister taxa to the group carried by *Myodes* voles. If the theory of hantavirus-host co-evolution is to be supported, this suggests that although host switching events definitely occurred, it is not certain which rodent taxa was the original host (Plyusnina et al. 2008a). None of the viruses have been linked to human disease but the pathogenicity of Topografov virus has not been completely ruled out (Vapalahti et al. 1999a, b).

9.3.3 Sigmodontinae and Neotominae-Borne Viruses

Sin Nombre virus was the first sigmodontine-hosted hantavirus to be identified during a outbreak of a previously unknown human disease in 1993, known as the 'Four Corners' outbreak named after the geographic location in which it occurred (MMWR 1993). This disease is now known as Hantavirus cardio-pulmonary syndrome and SNV is still the species most commonly associated with this syndrome in North America (Khan et al. 1996b; Rivers et al. 2009). SNV is hosted by the deer mouse, *Peromyscus maniculatus,* a generalist rodent species found widespread across North America. Worryingly evidence suggests that SNV prevalence increases with decreased rodent species diversity within the rodent community (Clay et al. 2009; Dizney and Ruedas 2009). Two variants, New York virus and Monogahela virus, hosted by *Peromyscus leucopus* and *P. maniculatus* respectively, are also included within this species (Fig. 9.5) and are also associated with human disease (Hjelle et al. 1995b; Song et al. 1994). Blue River virus identified in Oklahoma and Indiana, may also fall within this species complex (sequence data not shown). Although human cases are not numerous, outbreaks of disease caused by SNV do appear to be related to El Nino events in the south western USA and the corresponding effect on vegetation and rodent abundance (Hjelle and Glass 2000).

Rio Segundo virus was initially identified in 1994 and has been detected in several rodent species including *Peromyscus mexicanus* and several species in the *Reithrodontomys* genus (*R. mexicanus, R. sumichrasti, R. gracilis* and *R. creper*), suggesting that as with TULV, PUUV and ANDV, this virus species is hosted by more than one *Reithrodontomys* rodent species (Hjelle et al. 1995a; Salazar-Bravo et al. 2004). Closely related variants include El Moro Canyon virus, found in *R. megalotis* (Hjelle et al. 1994) and Limestone Canyon virus, detected in *Peromyscus boylii* apparently as a result of a host switch (Sanchez et al. 2001). Collectively, this

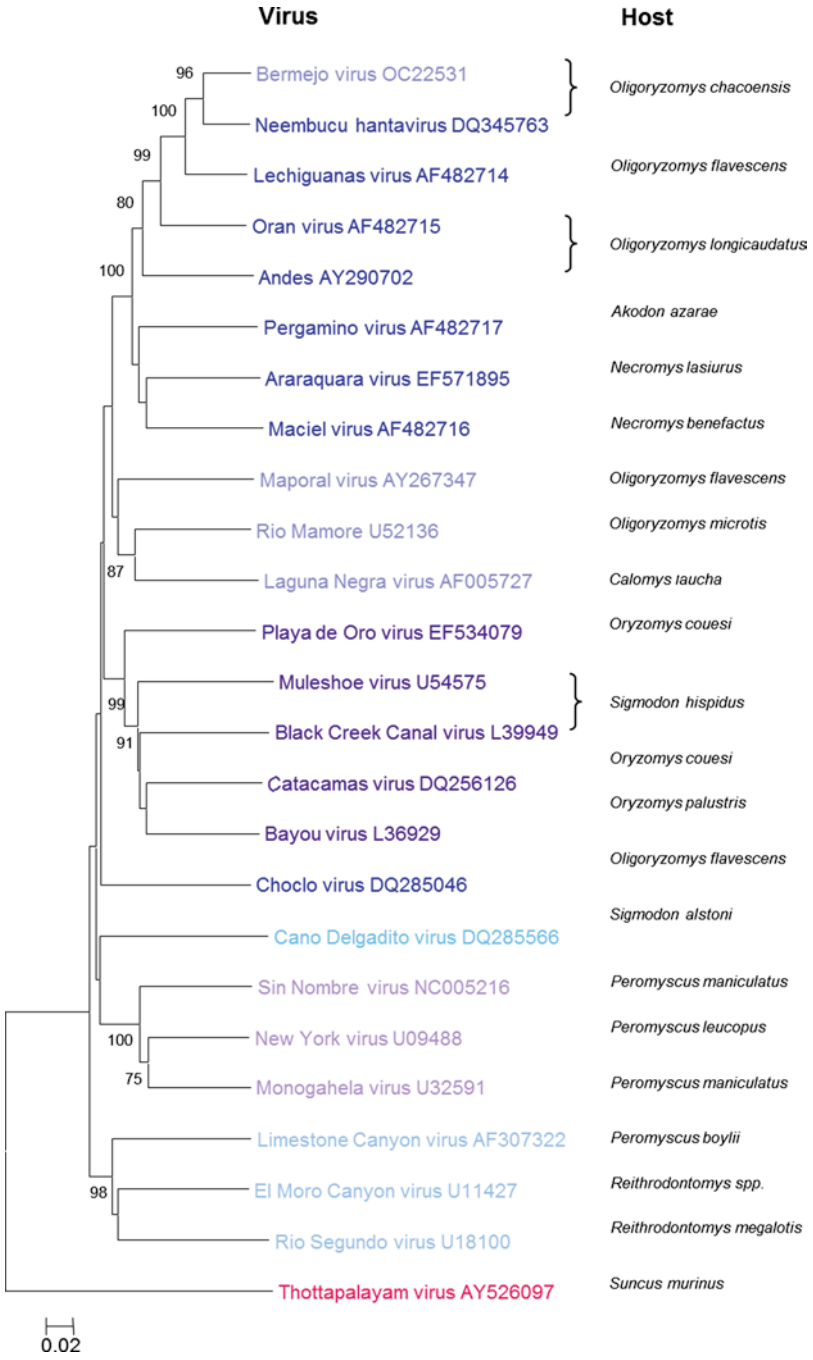

Fig. 9.5 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial S segment, showing bootstrap values ≥70 for representatives of all sigmodontine/neotomine-borne hantavirus species and rooted to Thottapalayam virus. GenBank accession numbers and proposed primary rodent hosts are shown next to each sequence. Branch lengths are to scale

virus has been identified in Costa Rica, Mexico, the southern USA and Panama. None of these lineages have been associated with human disease.

Bayou virus (BAYV) was the second North American hantavirus to be identified as a human pathogen in Louisiana in 1995 (Morzunov et al. 1995). It is hosted by the rice rat *Oryzomys palustris* (Torrez-Martinez and Hjelle 1995) and along with Black Creek Canal virus (isolated from *Sigmodon hispidus* and also identified in 1995 (Rollin et al. 1995)) with which it groups according to a recent revision of hantavirus taxonomy, it is responsible for a handful of human HCPS cases (Khan et al. 1996a; Torrez-Martinez et al. 1998). Muleshoe virus, also hosted by *Sigmodon hispidus* (Rawlings et al. 1996) is also placed within this species complex, but this lineage has not yet been associated with human disease. Two other viruses recently identified, Playa de Oro virus in *Oryzomys couesi* and *Sigmodon mascotensis* (Chu et al. 2008) from Mexico and Catacamas virus also in *Oryzomys couesi* but from Honduras (Milazzo et al. 2006), may also belong to this species (Fig. 9.5). Collectively these viruses appear to be restricted to the south-eastern USA and possibly Mexico (Chu et al. 2008).

Andes virus (ANDV), originally identified in 1995 (Lopez et al. 1996), is the major cause of HCPS in humans throughout much of South America. CFR is comparable to that observed for SNV (Campos et al. 2009; Castillo et al. 2001). Unlike all other hantaviruses, this species is the only species found to be capable of human-to-human transmission (Martinez et al. 2005). ANDV is actually comprised of many different lineages (Fig. 9.5), with each lineage found in different geographic regions. Lineages identified to date include Araraquara (Raboni et al. 2009), Choclo (Vincent et al. 2000), Lechiguanas (Levis et al. 1998), Maciel (Bohlman et al. 2002), Neembucu (Chu et al. 2003), Oran (Levis et al. 1998) and Pergamino viruses (Bohlman et al. 2002a). Four hantaviruses, previously proposed as separate species, namely Bermejo (Padula et al. 2002a), Maporal (Bohlman et al. 2002), Rio Mamore (Bharadwaj et al. 1997) and Laguna Negra (Johnson et al. 1997) viruses, may also be part of the Andes species-complex. Many of the ANDV lineages are hosted by rodents in the genus *Oligoryzomys*, but some of them are found in rodents from different genera, e.g. Laguna Negra virus in rodents of the genus *Calomys* (Johnson et al. 1997; Levis et al. 2004), Maciel virus in *Necromys benefactus* and Pergamino virus in the rodent species *Akodon azarae* (Bohlman et al. 2002).

Cano Delgadito virus was identified in a cotton rat, *Sigmodon alstoni* from Venezuela in 1997 (Fulhorst et al. 1997). Little work has been carried out on this virus, other than some experimental infection work on the natural host (Fulhorst et al. 2002) and the virus does not appear to be associated with human disease.

9.3.4 Soricomorph-Borne Viruses

Thottapalayam virus (TPMV) was the first hantavirus to be isolated and was identified in an Asian house shrew, *Suncus murinus*, in India in 1971 (Carey et al. 1971). This shrew species is widespread in both south and east Asia (Manuel et al. 1996), although the virus has only been recorded in this species in India and Indonesia

(Okumura et al. 2007). However, a serological study identified anti-TPMV antibodies in a Thai patient (Okumura et al. 2007). Little interest was shown in this virus species until recently, when hantaviruses were identified in several other insectivore species.

The second soricomorph-borne virus to be reported was Tanganya virus in March 2007. This virus was identified in *Crocidura theresea* in Guinea (Klempa et al. 2007), but it has not yet been fully characterised. When this virus was identified, the hypothesis of co-evolution between virus and host species was fairly universally accepted for hantaviruses, supported by the extreme genetic difference between TPMV and all other discovered hantaviruses. However, unexpectedly Tanganya virus was found to show the lowest sequence homology (for a section of the S segment) to TPMV (47.5% nucleotide and 39.4% amino acid identities) and the highest to Sangassou virus (62.1% nucleotide identity), the hantavirus identified in an African woodmouse from Guinea (Klempa et al. 2006, 2009). The co-evolution hypothesis was called further into question when several other soricomorph-borne hantaviruses were identified and analysed in quick succession.

Camp Ripley virus was identified in the Northern short-tailed shrew, *Blarina brevicauda*, in the USA in 2007 (Arai et al. 2007). This virus also groups more closely with rodent-borne hantaviruses than with TPMV, but sequence identities are lower than those shown by Tanganya virus (Kang et al. 2009c).

Seewis virus was detected in Switzerland in 2006 in the European common shrew, *Sorex araneus*, a species found widespread throughout Europe and into Central Siberia (Meinig et al. 2006). This virus also appears to group more closely with rodent-borne hantaviruses and especially with Ash River virus when the S segment is compared but with Asama virus when the M segment is considered (Kang et al. 2009b). It has since been detected in archived shrew tissues from both Finland, Hungary (Kang et al. 2009a) and Russia (Yashina et al. 2010) and related strains in the closely related shrew species, *S. caecutiens* in Siberia and Finland and *S minutus* in Finland (personal observations).

Cao Bang virus was identified by molecular methods in three Chinese mole shrews, *Anourosorex squamipes* in Vietnam in 2006 (Song et al. 2007c). This virus species is also only distantly related to Thottapalayam virus and appears to be most closely related to Asama and Jemez Springs viruses (Kang et al. 2009b).

The detection in the USA of Ash River virus, found in the masked shrew, *Sorex cinereus* and Jemez Springs virus in the Dusky shrew, *Sorex monticolus,* was reported in 2008 (Arai et al. 2008a). Several isolates of Jemez Spring virus were identified from several locations and this virus has since been found in other *Sorex* species. Ash River virus was found to group most closely with Seewis virus (Kang et al. 2009b).

Asama virus was the first hantavirus to be identified in the family Talpidae. This virus was found in Japan in the Japanese shrew mole, *Urotriches talpoides*, in 2008 (Arai et al. 2008b).

Oxbow virus, the second talpid-associated hantavirus to be identified, was found in the USA in an American shrew mole, *Neurotrichus gibbsii*, captured in 2003, but only published in 2009 (Kang et al. 2009b). This virus is most closely related to Jemez Springs virus, with 70–73% sequence homology.

With the exception of TPMV, all of these hantaviruses have been found to group phylogenetically with Tanganya virus and are more closely related to rodent-borne hantaviruses than to TPMV.

A further two soricomorph-borne viruses have recently been discovered. Imjin virus, identified in the Ussuri white-toothed shrew, *Crocidura lasiura*, in Korea was reported 2009. From the two isolates analysed, the S segment was found to differ from TPMV by 31.3% and 30.1% at the nucleotide and amino acid level respectively (Song et al. 2009). The same year, Nova virus was found in the European common mole, *Talpa europaea*, in Hungary (Kang et al. 2009c). These two species, along with TPMV have been identified as the most highly divergent hantaviruses and form a well supported out-group (Fig. 9.6), showing low sequence homologies to other hantaviruses in the range of 54–58% at the nucleotide level for the entire S segment and 60–64% at the nucleotide level for the entire L segment. Due to the highly divergent nature of Nova virus, the M segment has not yet been successfully amplified (Kang et al. 2009c).

Finally, Qiandao Lake virus has recently been identified in *Suncus cylindricauda* in Yunnan province in southern China (2010). The discovery of this virus has not yet been officially reported, but the full S and M segment sequences are available in public databases (Fig. 9.6). It is highly probable that a number of Soricomorpha-carried hantaviruses will be found in the near future.

9.4 Recent Breakthroughs

9.4.1 Geographic Distribution

In recent years, hantaviruses infections have been identified in several new host species. At least 14 new species (or lineages) have been discovered within the last 5 years, 11 from Soricid hosts (Tanganya, Camp Ripley, Cao Bang, Ash River, Jemez Springs, Asama, Oxbow, Imjin, Nova and Qiandao Lake viruses), 2 from murine hosts (Serang, including Jurong virus and Sangassou virus) and 1 from a Sigmodontine host (Catacamas virus, possibly a lineage of BAYV). Several other species from Soricid hosts have also been detected (three in the USA, two in Russia, two in Finland and one in Korea) but have not yet been reported officially (Kang et al. 2009c). The identification of these viruses has also contributed to a more accurate knowledge of hantaviruses distribution (Fig. 9.7).

Molecular methods are also providing more information in locations where human hantavirus infections are known but the causative hantavirus species is undetermined. Although TPMV was discovered in India long ago (Carey et al. 1971), it has only been in recent years that human infections have been observed (Clement et al. 2006; Gadkari 2005). According to molecular analysis, the virus (or viruses) responsible appears to be a murid-borne virus (Chandy et al. 2009b). A similar story has occurred in Africa. Although serological surveys detected antibodies to hantaviruses in human

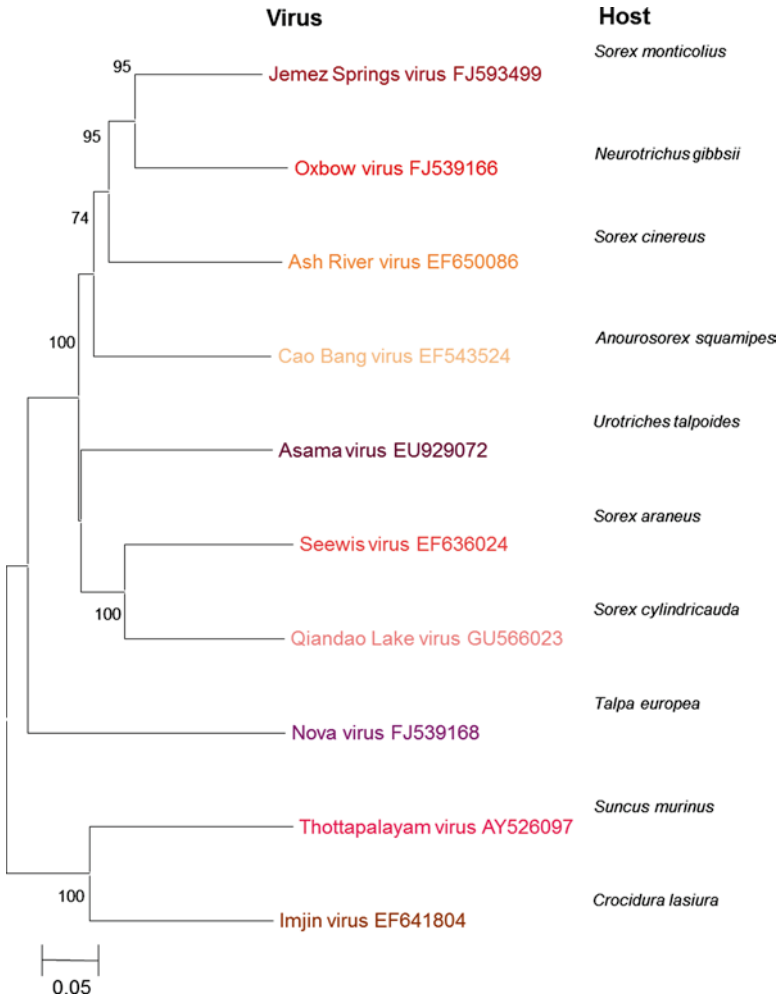

Fig. 9.6 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial S segment, showing bootstrap values ≥70 for representatives of all soricomorph-borne hantavirus species. GenBank accession numbers and proposed primary soricomorph hosts are shown next to each sequence. Branch lengths are to scale

populations in several west and central African countries (Gonzalez et al. 1984, 1989; Tomori et al. 1986), in Djibouti (Rodier et al. 1993) and in Egypt (Baddour et al. 1996), the first molecular evidence of hantavirus presence in sub-Saharan Africa was recorded only recently. In Guinea, Sangassou virus (from African woodmice) and Tanganya virus (from a shrew, *Crocidura theresae*) were identified using molecular techniques in 2006 and 2007 respectively (Klempa et al. 2006, 2007). As yet, neither virus has been implicated in human disease although a recent study in Sangassou village has detected antibodies in humans (Klempa et al. 2010).

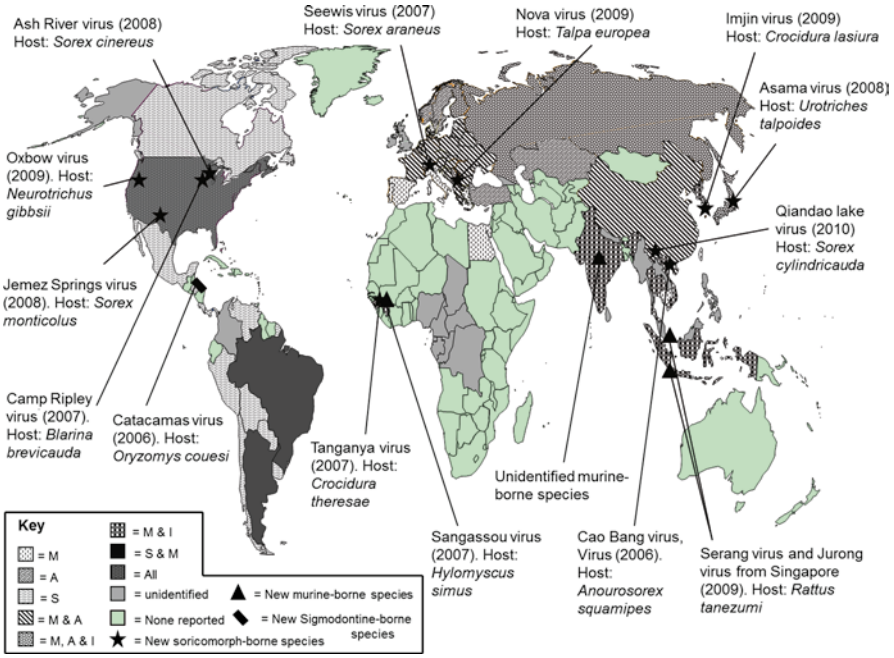

Fig. 9.7 Geographic distribution and hosts of recently discovered hantaviruses. Key: *M* urine-borne, *A* arvicoline-borne, *S* sigmodontine-borne, *I* soricomorph-borne

9.4.2 Hantavirus Hosts

9.4.2.1 Co-evolution vs. Host Switching

Until recently, the predominant view of hantavirus phylogeny was that these viruses had, in general, co-evolved with their rodent hosts. The highly divergent nature of TPMV, added further support to this theory (Song et al. 2007a). The co-evolution theory appeared to fit well both on family and subfamily level (Cricetidae versus Muridae, and within the former, Arvicolinae versus Sigmodontinea/Neotominae) (Herbreteau et al. 2006) and on a group and species level. For example the different DOBV-SAAV lineages are hosted by different rodent lineages of the genus *Apodemus* (Klempa et al. 2008) and the SNV variants are detected in different *Peromyscus* species (Monroe et al. 1999). Some well-quoted examples of host-switching events include the Arvicoline-borne virus, Topografov virus, which although hosted by *Lemmus sibiricus*, from co-phylogeny analysis, should be hosted by a member of the *Microtus* genus (Vapalahti et al. 1999a, b). Also the Sigmodontine associated Limestone Canyon virus, which would appear to be hosted by a rodent of the *Reithrodontomys* genus, when in fact it is found in *Permoyscus boylii* (Sanchez et al. 2001) (Fig. 9.8). However in general the co-evolution hypothesis based on rodent hantaviruses was fairly well accepted.

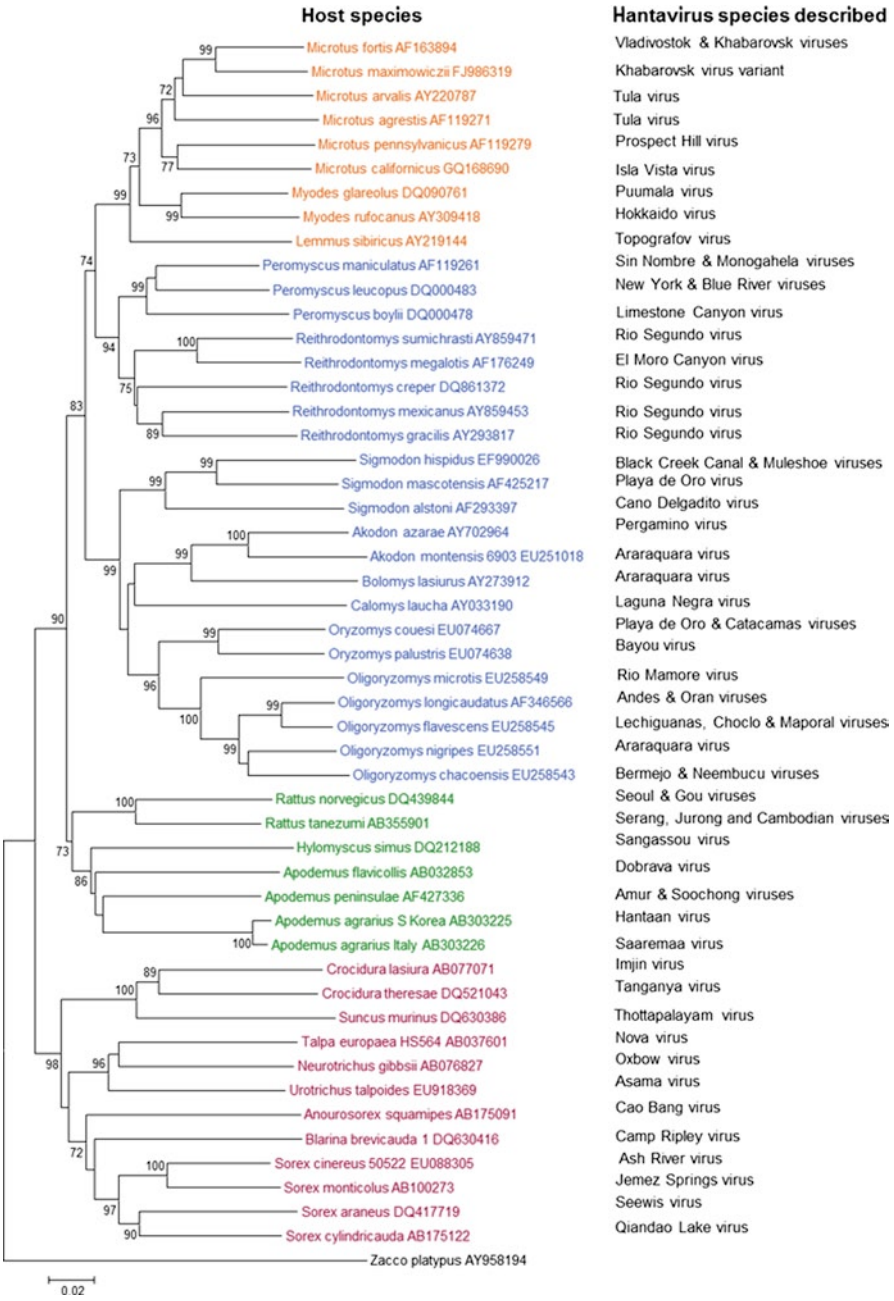

Fig. 9.8 Neighbour-joining tree using the p-distance model (10,000 replicates) for the partial cytochrome B gene of all rodent and soricomorph species suspected as primary hantavirus reservoirs. Tree is rooted to the cytochrome B sequence of the fish, *Zacco platypus*. GenBank accession numbers for host species and suggested associated hantaviruses are shown next to each sequence. Branch lengths are to scale

However, the recent discovery of several species of hantaviruses hosted by soricomorph mammals, suggested that this co-evolution picture may not be quite so simple (Ramsden et al. 2007). If the co-evolution theory was to stand, then these viruses, hosted by much more distantly related mammals, a totally different order (Fig. 9.8) should in theory show a much higher divergence to the rodent-borne hantaviruses. Although this is the case for Imjin, Thottapalayam and Nova viruses, it clearly is not for the remaining soricomorph-borne species (e.g. Seewis virus, Asama virus, Ash River virus, etc.), all of which group much more closely with the rodent-borne viruses than they do with the other soricomorph-borne viruses. There has even been some evidence of co-evolution and host-switching within this latter group (Kang et al. 2009b), but it is clear that other explanations are needed to explain what we currently see in terms of hantavirus genetic evolution (Ramsden et al. 2008).

Three soricomorph hantaviruses (namely: Thottapalayam, Imjin and Nova viruses) appear to be the most genetically divergent of the *Hantavirus* genus. This suggests that soricomorphs share the longest evolutionary history with hantaviruses and that ancestral soricomorphs may have been the ancestral hosts of hantaviruses. Viruses of the *Hantavirus* genus are the only Bunyaviridae not to require arthropod vectors. The possibility that insectivorous mammals may have been the original hosts of the hantaviruses suggests that an opportunity may have arisen that allowed an originally insect-borne virus, to adapt and then evolve in mammalian hosts, no longer requiring arthropod vectors for its transmission (Henttonen et al. 2008). Some hantaviruses have actually been detected in arthropod vectors, including HNTV in the trombiculid mite *Leptotrombidium scutellare* (Deng et al. 2008) and gamasid mites in China (Zhang et al. 2002) and BAYV in trombiculid mites and in an ixodid tick in the USA (Houck et al. 2001), although this does not imply that they have any role in hantavirus epidemiology.

Although rodents and soricomorphs are from different taxonomic orders, they often share the same habitat, potentially providing multiple opportunities for interaction. As hantaviruses have been shown to be fairly stable *ex vivo*, this may also build the case for environmental contamination (Hardestam et al. 2007). The presence of hantaviruses in both host orders suggests that some of these interactions (e.g. shared use of trails, inter-specific wounding, etc.) may have led to historical host-switching events. The existence of two genetically distinct groups of hantaviruses in soricomorphs indicates that the evolutionary history of hantaviruses is probably more complex than first thought. Perhaps hantaviruses originally infected soricomorphs, co-evolved to some extent in this order, then switched to rodents, co-evolving in this order and then switched back (perhaps on multiple occasions) to soricomorphs (Kang et al. 2009c). Hopefully the further discovery of more hantaviruses in other soricomorphs (and rodents) may help to provide further clarification of this situation.

9.4.2.2 Primary Hosts Species Identification

Mammalogists acknowledged for a long that the taxonomy of rodent species needed some revision. Many rodent species have been given several colloquial and even

scientific names over the years. Voucher specimens or information of great importance (i.e., collection locality) are still often missing. The precise identification of the host is crucial to study presumed co-evolution of the animal with its virus. As stated previously, one of the defining features of each hantavirus is that it is hosted by a different host species or sub-species (Nichol et al. 2005). Molecular techniques are now regularly employed to identify new rodent and soricomorph specimens and to establish the identity of archival voucher specimens. Generally the techniques used include mitochondrial DNA analysis (e.g. cytochrome B gene or cytochrome oxidase subunit 1 sequences) or other barcode markers in the host genome and this has led to the identification of cryptic species and a revision of many of the rodent genera and species (Lecompte et al. 2005; Pages et al. 2010; Rivera et al. 2007). As a result, many studies now routinely genetically determine the rodent or soricomorph host species consecutively with the hantavirus species (Plyusnina et al. 2008b; Kang et al. 2009a; Korva et al. 2009a).

The molecular identification and analysis of both host and hantavirus species has added to the knowledge of hantavirus-host relationships. It is now clear that although many hantaviruses have a single dominant rodent host (or soricomorph) species reservoir, many regularly cause spill-over infections into other species. In addition, there are also several examples of host-sharing. In some cases, such as Rio Segundo virus, there does not appear to be any significant genetic differences between the virus strains infecting different host species or lineages (Nemirov et al. 2010; Salazar-Bravo et al. 2004). But in some other virus species (e.g., DOBV-SAAV and ANDV), different virus strains infect different rodent species or sub-species.

9.4.2.3 Host Phylogeography

Molecular studies of hantaviruses may also provide clues to the historical and current migratory pathways and dispersal routes of their small mammalian hosts. In some viruses, closely related lineages are found in different geographic locations (Dekonenko et al. 2003; Zou et al. 2008a), suggesting that there has been fairly recent transmission of hantaviruses between the mammalian hosts at these locations. At some sites, hantaviruses are sporadically present and absent, indicating re-infection from an outside source and therefore present migration and/or dispersal routes (Abramson and Kenkre 2002). The presence of genetically different lineages at different geographic locations suggests an absence of gene flow between these locations, perhaps indicating geographic isolation or alternatively, providing clues to historical refugia. Areas of high hantavirus prevalence may also be linked to current or historical refugia (Dragoo et al. 2006). With SNV, human disease was more frequent in areas where the deer mouse population was persistently infected, in contrast to areas where the deer mouse population was only intermittently infected (Feuer et al. 1999).

The presence of HNTV and SAAV in eastern and western lineages of *Apodemus agrarius* (Plyusnina et al. 2009a; Zou et al. 2008a) also raises some interesting questions. SAAV is closer to DOBV than to HNTV, but the split between western

and eastern *A. agrarius* in no more than 100 000 years old (Michaux, personal communication). It may be that as yet undetected intermediate virus forms exist in other *Apodemus* species, potentially in central Asia or China.

9.4.2.4 Host Species Susceptibility

Although the shared genetic variability between host and virus may be explained to some extent by a shared evolutionary history (such as migration, isolation and refugia), a kind of 'arms race' may also exist between hantavirus and host, leading to further genetic variation. Indeed, in some areas that do support the appropriate hantavirus host species, hantaviruses are not present. Although this may be due to historical reasons, variation in rodent major histocompatibility complex (MHC) subtypes that relate to susceptibility to infection by hantaviruses have also been described and may play a role in hantavirus geographic range. At least three MHC alleles in bank voles (*Myodes glareolus*) affect the individual's infection susceptibility. Allele Cgl-DQA-09 is positively associated with infection with PUUV, while alleles Cgl-DQA-05 and Cgl-DQA-12 reduce this risk (Deter et al. 2008). It also appears that some rodent species, although present at large enough densities to support hantavirus circulation, have never been found to host hantaviruses, suggesting possible genetic resistance to hantavirus infection in these animals (N. Charbonnel, personal communication).

9.4.3 Recombination and Reassortment

Hantaviruses, as RNA viruses, show a fairly high degree of genetic variability. The presence of quasi-species has been demonstrated both in the laboratory and under natural conditions for several hantaviruses (Feuer et al. 1999; Plyusnin et al. 1996a; Sironen et al. 2008). The presence of several viral sub-populations or even virus species within the same host individual provides opportunities for adaptation to a potentially changing environment and for both re-assortment and recombination to occur.

Laboratory induced and natural reassortment and recombination of hantavirus strains has been detected for several hantaviruses. Non-random reassortment has been observed between PUUV strains in Finland at the contact zone where two PUUV lineages meet (Razzauti et al. 2009). S and L segments always came from the same lineage and this non-random pattern suggests that this genome configuration is important for virus survival. Natural reassortment from the same lineages have also been detected for PUUV (Razzauti et al. 2008), SNV (focusing primarily on gene-flow in the M segment) (Black et al. 2009), ANDV (Medina et al. 2009), HNTV (Zou et al. 2008a) and potentially historically in DOBV (Klempa et al. 2003b). Recombination events have also been demonstrated for the S segment of Tula hantavirus (Sibold et al. 1999). A natural reassortment virus between HTNV and SEOV has also been observed (Zou et al. 2008b), indicating that reassortment

viruses may occur between species as well as within species in the right environment, in contrary to one of the ICTV's rules regarding hantavirus species classification (Nichol et al. 2005).

9.4.4 Susceptibility to Disease in Humans

Recent studies have indicated that susceptibility to infection and disease severity in humans could be associated with immunogenetic factors. Individuals with certain human leukocyte antigen (HLA) haplotypes or tumour necrosis factor polymorphisms are more predisposed to severe disease. For example HLA haplotype B8-DR3, C4A*Q0 and DRB1*0301 individuals are at risk of developing severe Puumala disease (Makela et al. 2002; Mustonen et al. 1996). Populations with a higher occurrence of certain haplotypes (e.g. HLA-B*46-DRB1*09 and HLA-DRB1*09 in the Chinese Han population) also seem to be more susceptible to HNTV infection (Wang et al. 2009). In patients infected with ANDV, HLA-DRB 1*15 was commonly associated with a mild disease course, while HLA-B*08 was more common in patients with severe disease (Ferrer et al. 2007). Patients with either increased or decreased levels of tumour necrosis factor alpha are also more predisposed to experience severe disease during Puumala infection (Kanerva et al. 1998; Maes et al. 2006; Makela et al. 2002). The platelet glycoprotein polymorphism, HPA-3, also appears to be linked to more severe disease course in humans (Liu et al. 2009). This may help to explain the variation in disease course seen in many patients and the geographic variability in disease incidence, although as mentioned previously, some areas with a high incidence of human hantavirus infection, may relate to viral refugia sites (Glass et al. 2007).

As with rodents, the susceptibility to hantavirus disease in humans varies with gender. Although similar proportions of men and women are known to have antibodies to PUUV, men are more likely to express severe symptoms (Vapalahti et al. 2003). However, when assessing disease risk for different individuals, other non-genetic risk factors generally relating to hantavirus exposure such as an individual's occupations or activities, must be considered. Individuals engaged in outdoor sports or occupations (e.g. forestry or farming) or in peridomestic cleaning are at higher risk of hantavirus infection in some locations (Pal et al. 2005; Vapalahti et al. 1999a, b; Zeitz et al. 1995), but not in all (Fulhorst et al. 2007; Schultze et al. 2007). Smoking is also a clear risk factor for hantavirus infection (Vapalahti et al. 2010).

9.5 Conclusion

During recent years, molecular methods have significantly contributed to the comprehension of hantavirus biology and diversity, the role of reservoir hosts and the importance of the human disease. A wider use of the existing methods and the development of new molecular tools will hopefully provide more precise data and address some of the important questions still under much debate.

References

Abramson G, Kenkre VM (2002) Spatiotemporal patterns in the Hantavirus infection. Phys Rev E Stat Nonlin Soft Matter Phys 66(011912):1–5

Aplin KP, Brown PR, Jacob J et al (2003) Field methods for rodent studies in Asia and the Indo-Pacific. ACIAR, Melbourne

Arai S, Song JW, Sumibcay L et al (2007) Hantavirus in northern short-tailed shrew, United States. Emerg Infect Dis 13:1420–1423

Arai S, Bennett SN, Sumibcay L et al (2008a) Phylogenetically distinct hantaviruses in the masked shrew (Sorex cinereus) and dusky shrew (Sorex monticolus) in the United States. Am J Trop Med Hyg 78:348–351

Arai S, Ohdachi SD, Asakawa M et al (2008b) Molecular phylogeny of a newfound hantavirus in the Japanese shrew mole (Urotrichus talpoides). Proc Natl Acad Sci USA 105:16296–16301

Arthur RR, Lofts RS, Gomez J et al (1992) Grouping of hantaviruses by small (S) genome segment polymerase chain reaction and amplification of viral RNA from wild-caught rats. Am J Trop Med Hyg 47:210–224

Asada H, Balachandra K, Tamura M et al (1989) Cross-reactive immunity among different serotypes of virus causing hemorrhagic fever with renal syndrome. J Gen Virol 70(Pt 4):819–825

Avsic-Zupanc T, Xiao SY, Stojanovic R et al (1992) Characterization of Dobrava virus: a Hantavirus from Slovenia, Yugoslavia. J Med Virol 38:132–137

Baddour MM, Mourad AS, Awwad AM et al (1996) Hantaan virus infection among human and rat populations in Alexandria. J Egypt Public Health Assoc 71:213–228

Bharadwaj M, Botten J, Torrez-Martinez N et al (1997) Rio Mamore virus: genetic characterization of a newly recognized hantavirus of the pygmy rice rat, Oligoryzomys microtis, from Bolivia. Am J Trop Med Hyg 57:368–374

Bi Z, Formenty PB, Roth CE (2008) Hantavirus infection: a review and global update. J Infect Dev Ctries 2:3–23

Black WCt, Doty JB, Hughes MT et al (2009) Temporal and geographic evidence for evolution of Sin Nombre virus using molecular analyses of viral RNA from Colorado, New Mexico and Montana. Virol J 6:102

Bohlman MC, Morzunov SP, Meissner J et al (2002) Analysis of hantavirus genetic diversity in Argentina: S segment-derived phylogeny. J Virol 76:3765–3773

Brummer-Korvenkontio M, Vaheri A, Hovi T et al (1980) Nephropathia epidemica: detection of antigen in bank voles and serologic diagnosis of human infection. J Infect Dis 141:131–134

Campos GM, Borges AA, Badra SJ et al (2009) Pulmonary and cardiovascular syndrome due to hantavirus: clinical aspects of an emerging disease in southeastern Brazil. Rev Soc Bras Med Trop 42:282–289

Carey DE, Reuben R, Panicker KN et al (1971) Thottapalayam virus: a presumptive arbovirus isolated from a shrew in India. Indian J Med Res 59:1758–1760

Castillo C, Naranjo J, Sepulveda A et al (2001) Hantavirus pulmonary syndrome due to Andes virus in Temuco, Chile: clinical experience with 16 adults. Chest 120:548–554

Chan YC, Wong TW, Yap EH (1987) Hemorrhagic fever with renal syndrome: clinical, virological and epidemiological perspectives. Ann Acad Med Singapore 16:696–701

Chandy S, Abraham P, Sridharan G (2008) Hantaviruses: an emerging public health threat in India? A review. J Biosci 33:495–504

Chandy S, Okumura M, Yoshimatsu K et al (2009a) Hantavirus species in India: a retrospective study. Indian J Med Microbiol 27:348–350

Chandy S, Yoshimatsu K, Boorugu HK et al (2009b) Acute febrile illness caused by hantavirus: serological and molecular evidence from India. Trans R Soc Trop Med Hyg 103:407–412

Chen HX, Qiu FX, Dong BJ et al (1986) Epidemiological studies on hemorrhagic fever with renal syndrome in China. J Infect Dis 154:394–398

Chu YK, Owen RD, Gonzalez LM et al (2003) The complex ecology of hantavirus in Paraguay. Am J Trop Med Hyg 69:263–268

Chu YK, Owen RD, Sanchez-Hernandez C et al (2008) Genetic characterization and phylogeny of a hantavirus from Western Mexico. Virus Res 131:180–188
Chumakov MP, Gavrilovskaya IN, Boiko VA et al (1981) Detection of hemorrhagic fever with renal syndrome (HFRS) virus in the lungs of bank voles Clethrionomys glareolus and redbacked voles Clethrionomys rutilus trapped in HFRS foci in the European part of U.S.S.R., and serodiagnosis of this infection in man. Arch Virol 69:295–300
Clay CA, Lehmer EM, Jeor SS et al (2009) Sin nombre virus and rodent species diversity: a test of the dilution and amplification hypotheses. PLoS One 4:e6467
Clement J, Maes P, Muthusethupathi M et al (2006) First evidence of fatal hantavirus nephropathy in India, mimicking leptospirosis Nephrol Dial. Transplant 21:826–827
Clement J, Vercauteren J, Verstraeten WW et al (2009) Relating increasing hantavirus incidences to the changing climate: the mast connection. Int J Health Geogr 8:1
Coulaud X, Chouaib E, Georges AJ et al (1987) First human case of hemorrhagic fever with renal syndrome in the Central African Republic. Trans R Soc Trop Med Hyg 81:686
Cueto GR, Cavia R, Cet B, Bellomo C et al (2008) Prevalence of hantavirus infection in wild Rattus norvegicus and R. rattus populations of Buenos Aires City, Argentina. Trop Med Int Health 13:46–51
Daud NH, Kariwa H, Tanikawa Y et al (2007) Mode of infection of Hokkaido virus (Genus Hantavirus) among grey red-backed voles, Myodes rufocanus, in Hokkaido, Japan. Microbiol Immunol 51:1081–1090
Dekonenko A, Yakimenko V, Ivanov A et al (2003) Genetic similarity of Puumala viruses found in Finland and western Siberia and of the mitochondrial DNA of their rodent hosts suggests a common evolutionary origin. Infect Genet Evol 3:245–257
Deng XZ, Zhang Y, Kong J et al (2008) Molecular epidemiological study on the host and role of the Hantavirus and Orientia tsutsugamushi in the same epidemic area. Zhonghua Yu Fang Yi Xue Za Zhi 42:574–577
Deter J, Bryja J, Chaval Y et al (2008) Association between the DQA MHC class II gene and Puumala virus infection in Myodes glareolus, the bank vole. Infect Genet Evol 8:450–458
Dixon B (2009) Beware fat bank voles. Lancet Infect Dis 9:210
Dizney LJ, Ruedas LA (2009) Increased host species diversity and decreased prevalence of Sin Nombre virus. Emerg Infect Dis 15:1012–1018
Dragoo JW, Lackey JA, Moore KE et al (2006) Phylogeography of the deer mouse (Peromyscus maniculatus) provides a predictive framework for research on hantaviruses. J Gen Virol 87:1997–2003
Easterbrook JD, Kaplan JB, Vanasco NB et al (2007) A survey of zoonotic pathogens carried by Norway rats in Baltimore, Maryland, USA. Epidemiol Infect 135:1192–1199
Elliott LH, Ksiazek TG, Rollin PE et al (1994) Isolation of the causative agent of hantavirus pulmonary syndrome. Am J Trop Med Hyg 51:102–108
Escutenaire S, Chalon P, De Jaegere F et al (2002) Behavioral, physiologic, and habitat influences on the dynamics of Puumala virus infection in bank voles (Clethrionomys glareolus). Emerg Infect Dis 8:930–936
Ferrer JF, Galligan D, Esteban E et al (2003) Hantavirus infection in people inhabiting a highly endemic region of the Gran Chaco territory, Paraguay: association with Trypanosoma cruzi infection, epidemiological features and haematological characteristics. Ann Trop Med Parasitol 97:269–280
Ferrer CP, Vial CPA, Ferrés GM et al (2007) Genetic susceptibility to Andes Hantavirus: association between severity of disease and HLA alleles in Chilean patients. Rev Chilena Infectol 24:351–359
Feuer R, Boone JD, Netski D et al (1999) Temporal and spatial analysis of Sin Nombre virus quasispecies in naturally infected rodents. J Virol 73:9544–9554
Forthal DN, Bauer SP, McCormick JB (1987) Antibody to hemorrhagic fever with renal syndrome viruses (Hantaviruses) in the United States. Am J Epidemiol 126:1210–1213

Fulhorst CF, Monroe MC, Salas RA et al (1997) Isolation, characterization and geographic distribution of Cano Delgadito virus, a newly discovered South American hantavirus (family Bunyaviridae). Virus Res 51:159–171

Fulhorst CF, Milazzo ML, Duno G et al (2002) Experimental infection of the Sigmodon alstoni cotton rat with Cano Delgadito virus, a South American hantavirus. Am J Trop Med Hyg 67:107–111

Fulhorst CF, Milazzo ML, Armstrong LR et al (2007) Hantavirus and arenavirus antibodies in persons with occupational rodent exposure. Emerg Infect Dis 13:532–538

Gadkari DA (2005) Hantaviruses are here in India. Indian J Med Res 122:193–195

Glass GE, Childs JE, Korch GW et al (1988) Association of intraspecific wounding with hantaviral infection in wild rats (Rattus norvegicus). Epidemiol Infect 101:459–472

Glass GE, Watson AJ, LeDuc JW et al (1994) Domestic cases of hemorrhagic fever with renal syndrome in the United States. Nephron 68:48–51

Glass GE, Shields T, Cai B et al (2007) Persistently highest risk areas for hantavirus pulmonary syndrome: potential sites for refugia. Ecol Appl 17:129–139

Golovljova I, Sjolander KB, Lindegren G et al (2002) Hantaviruses in Estonia. J Med Virol 68:589–598

Golovljova I, Vasilenko V, Mittzenkov V et al (2007) Characterization of hemorrhagic fever with renal syndrome caused by hantaviruses, Estonia. Emerg Infect Dis 13:1773–1776

Gonzalez JP, McCormick JB, Baudon D et al (1984) Serological evidence for Hantaan-related virus in Africa. Lancet 2:1036–1037

Gonzalez JP, Josse R, Johnson ED et al (1989) Antibody prevalence against hemorrhagic fever viruses in randomized representative Central African populations. Res Virol 140:319–331

Hardestam J, Simon M, Hedlund KO et al (2007) Ex vivo stability of the rodent-borne Hantaan virus in comparison to that of arthropod-borne members of the Bunyaviridae family. Appl Environ Microbiol 73:2547–2551

Henttonen H, Buchy P, Suputtamongkol Y et al (2008) Recent discoveries of new hantaviruses widen their range and question their origins. Ann N Y Acad Sci 1149:84–89

Herbreteau V, Gonzalez JP, Hugot JP (2006) Implication of phylogenetic systematics of rodent-borne hantaviruses allows understanding of their distribution. Ann N Y Acad Sci 1081:39–56

Heyman P, Vaheri A (2008) Situation of hantavirus infections and hemorrhagic fever with renal syndrome in European countries as of December 2006. Euro Surveill 13:pii: 18925

Heyman P, Klingstrom J, de Jaegere F et al (2002) Tula hantavirus in Belgium. Epidemiol Infect 128:251–256

Heyman P, Plyusnina A, Berny P et al (2004) Seoul hantavirus in Europe: first demonstration of the virus genome in wild Rattus norvegicus captured in France. Eur J Clin Microbiol Infect Dis 23:711–717

Heyman P, Baert K, Plyusnina A et al (2009) Serological and genetic evidence for the presence of Seoul hantavirus in Rattus norvegicus in Flanders, Belgium. Scand J Infect Dis 41:51–56

Hjelle B, Glass GE (2000) Outbreak of hantavirus infection in the four corners region of the United States in the wake of the 1997–1998 El Nino-southern oscillation. J Infect Dis 181:1569–1573

Hjelle B, Chavez-Giles F, Torrez-Martinez N et al (1994) Genetic identification of a novel hantavirus of the harvest mouse Reithrodontomys megalotis. J Virol 68:6751–6754

Hjelle B, Anderson B, Torrez-Martinez N et al (1995a) Prevalence and geographic genetic variation of hantaviruses of New World harvest mice (Reithrodontomys): identification of a divergent genotype from a Costa Rican Reithrodontomys mexicanus. Virology 207:452–459

Hjelle B, Krolikowski J, Torrez-Martinez N et al (1995b) Phylogenetically distinct hantavirus implicated in a case of hantavirus pulmonary syndrome in the northeastern United States. J Med Virol 46:21–27

Horling J, Chizhikov V, Lundkvist A et al (1996) Khabarovsk virus: a phylogenetically and serologically distinct hantavirus isolated from Microtus fortis trapped in far-east Russia. J Gen Virol 77(Pt 4):687–694

Houck MA, Qin H, Roberts HR (2001) Hantavirus transmission: potential role of ectoparasites. Vector Borne Zoonotic Dis 1:75–79
Hugot JP, Plyusnina A, Herbreteau V et al (2006) Genetic analysis of Thailand hantavirus in Bandicota indica trapped in Thailand. Virol J 3:72
Hung T, Chou ZY, Zhao TX et al (1985) Morphology and morphogenesis of viruses of hemorrhagic fever with renal syndrome (HFRS). I. Some peculiar aspects of the morphogenesis of various strains of HFRS virus. Intervirology 23:97–108
Jaaskelainen KM, Kaukinen P, Minskaya ES et al (2007) Tula and Puumala hantavirus NSs ORFs are functional and the products inhibit activation of the interferon-beta promoter. J Med Virol 79:1527–1536
Jalbert Y (1997) Of mice and men: social impact of an emerging illness. Can J Public Health 88:134–136
Johansson P, Yap G, Low HT et al (2010) Molecular characterization of two hantavirus strains from different rattus species in Singapore. Virol J 7:15
Johnson AM, Bowen MD, Ksiazek TG et al (1997) Laguna Negra virus associated with HPS in western Paraguay and Bolivia. Virology 238:115–127
Kallio ER, Klingstrom J, Gustafsson E et al (2006) Prolonged survival of Puumala hantavirus outside the host: evidence for indirect transmission via the environment. J Gen Virol 87:2127–2134
Kanerva M, Vaheri A, Mustonen J et al (1998) High-producer allele of tumour necrosis factor-alpha is part of the susceptibility MHC haplotype in severe puumala virus-induced nephropathia epidemica. Scand J Infect Dis 30:532–534
Kang HJ, Arai S, Hope AG et al (2009a) Genetic diversity and phylogeography of Seewis virus in the Eurasian common shrew in Finland and Hungary. Virol J 6:208
Kang HJ, Bennett SN, Dizney L et al (2009b) Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (Neurotrichus gibbsii). Virology 388:8–14
Kang HJ, Bennett SN, Sumibcay L et al (2009c) Evolutionary insights from a genetically divergent hantavirus harbored by the European common mole (Talpa europaea). PLoS One 4:e6149
Kariwa H, Yoshimatsu K, Sawabe J et al (1999) Genetic diversities of hantaviruses among rodents in Hokkaido, Japan and Far East Russia. Virus Res 59:219–228
Khan AS, Gaviria M, Rollin PE et al (1996a) Hantavirus pulmonary syndrome in Florida: association with the newly identified Black Creek Canal virus. Am J Med 100:46–48
Khan AS, Khabbaz RF, Armstrong LR et al (1996b) Hantavirus pulmonary syndrome: the first 100 US cases. J Infect Dis 173:1297–1303
Kim YS, Ahn C, Han JS et al (1995) Hemorrhagic fever with renal syndrome caused by the Seoul virus. Nephron 71:419–427
Klempa B, Meisel H, Rath S et al (2003a) Occurrence of renal and pulmonary syndrome in a region of northeast Germany where Tula hantavirus circulates. J Clin Microbiol 41:4894–4897
Klempa B, Schmidt HA, Ulrich R et al (2003b) Genetic interaction between distinct Dobrava hantavirus subtypes in Apodemus agrarius and A. flavicollis in nature. J Virol 77:804–809
Klempa B, Fichet-Calvet E, Lecompte E et al (2006) Hantavirus in African wood mouse, Guinea. Emerg Infect Dis 12:838–840
Klempa B, Fichet-Calvet E, Lecompte E et al (2007) Novel hantavirus sequences in Shrew, Guinea. Emerg Infect Dis 13:520–522
Klempa B, Tkachenko EA, Dzagurova TK et al (2008) Hemorrhagic fever with renal syndrome caused by 2 lineages of Dobrava hantavirus, Russia. Emerg Infect Dis 14:617–625
Klempa B, Koivogui L, Sylla O et al (2010) Serological evidence of human hantavirus infections in Guinea, West Africa. J Infect Dis 201:1031–1034
Korva M, Duh D, Puterle A et al (2009a) First molecular evidence of Tula hantavirus in Microtus voles in Slovenia. Virus Res 144:318–322
Korva M, Duh D, Saksida A et al (2009b) The hantaviral load in tissues of naturally infected rodents. Microbes Infect 11:344–351
Kramski M, Meisel H, Klempa K et al (2007) Detection and typing of human pathogenic hantaviruses by real-time reverse transcription-PCR and pyrosequencing. Clin Chem 53:1899–1905

Kramski M, Achazi K, Klempa B et al (2009) Nephropathia epidemica with a 6-week incubation period after occupational exposure to Puumala hantavirus. J Clin Virol 44:99–101
Lecompte E, Brouat C, Duplantier J-M et al (2005) Molecular identification of four cryptic species of Mastomys (Rodentia, Murinae). Biochem Syst Ecol 33:681–689
Lee HW (1989) Hemorrhagic fever with renal syndrome in Korea. Rev Infect Dis 11(Suppl 4):S864–S876
Lee HW, Lee PW, Johnson KM (1978) Isolation of the etiological agent of Korean hemorrhagic fever. J Infect Dis 137:298–308
Lee PW, Amyx HL, Gajdusek DC et al (1982a) New hemorrhagic fever with renal syndrome-related virus in rodents in the United States. Lancet 2:1405
Lee HW, Baek LJ, Johnson KM (1982b) Isolation of Hantaan virus, the etiological agent of Korean hemorrhagic fever from wild urban rats. J Infect Dis 146:638–644
Lee PW, Amyx HL, Yanagihara R et al (1985) Partial characterization of Prospect Hill virus isolated from meadow voles in the United States. J Infect Dis 152:826–829
Levis S, Morzunov SP, Rowe JE et al (1998) Genetic diversity and epidemiology of hantaviruses in Argentina. J Infect Dis 177:529–538
Levis S, Garcia J, Pini N et al (2004) Hantavirus pulmonary syndrome in northwestern Argentina: circulation of Laguna Negra virus associated with Calomys callosus. Am J Trop Med Hyg 71:658–663
Linard C, Tersago K, Leirs H et al (2007) Environmental conditions and Puumala virus transmission in Belgium. Int J Health Geogr 6:55
Liu Z, Gao M, Han Q et al (2009) Platelet glycoprotein IIb/IIIa (HPA-1 and HPA-3) polymorphisms in patients with hemorrhagic fever with renal syndrome. Hum Immunol 70:452–456
Lokugamage K, Kariwa H, Hayasaka D et al (2002) Genetic characterization of hantaviruses transmitted by the Korean field mouse (Apodemus peninsulae), Far East Russia. Emerg Infect Dis 8:768–776
Lokugamage K, Kariwa H, Lokugamage N et al (2004) Genetic and antigenic characterization of the Amur virus associated with hemorrhagic fever with renal syndrome. Virus Res 101:127–134
Lopez N, Padula P, Rossi C et al (1996) Genetic identification of a new hantavirus causing severe pulmonary syndrome in Argentina. Virology 220:223–226
Maes P, Clement J, Groeneveld PH et al (2006) Tumor necrosis factor-alpha genetic predisposing factors can influence clinical severity in nephropathia epidemica. Viral Immunol 19:558–564
Maes P, Klempa B, Clement J et al (2009) A proposal for new criteria for the classification of hantaviruses, based on S and M segment protein sequences. Infect Genet Evol 9:813–820
Makela S, Mustonen J, Ala-Houhala I et al (2002) Human leukocyte antigen-B8-DR3 is a more important risk factor for severe Puumala hantavirus infection than the tumor necrosis factor-alpha(−308) G/A polymorphism. J Infect Dis 186:843–846
Manuel R, Christine C, Peter V et al (1996) Genetic differentiation and zoogeography of Asian Suncus murinus (Mammalia: Soricidae). Biol J Linn Soc 57:307–316
Martinez VP, Bellomo C, San Juan J et al (2005) Person-to-person transmission of Andes virus. Emerg Infect Dis 11:1848–1853
McCormick JB, Sasso DR, Palmer EL et al (1982) Morphological identification of the agent of Korean hemorrhagic fever (Hantaan virus)as a member of the Bunyaviridae. Lancet 1:765–768
Medina RA, Torres-Perez F, Galeno H et al (2009) Ecology, genetic diversity, and phylogeographic structure of andes virus in humans and rodents in Chile. J Virol 83:2446–2459
Meinig H, Kryštufek B, Sheftel B, et al. (2006) In: In IUCN 2010. IUCN red list of threatened species
Milazzo ML, Cajimat MN, Hanson JD et al (2006) Catacamas virus, a hantaviral species naturally associated with Oryzomys couesi (Coues' oryzomys) in Honduras. Am J Trop Med Hyg 75:1003–1010
Mills JN (2006) Biodiversity loss and emerging infectious disease: an example from the rodent-borne hemorrhagic fevers. Biodiversity 7:9–17
Mills JN, Yates TL, Ksiazek TG et al (1999) Long-term studies of hantavirus reservoir populations in the southwestern United States: rationale, potential, and methods. Emerg Infect Dis 5:95–101

MMWR (1993) Outbreak of acute illness–southwestern United States, 1993. Morb Mortal Wkly Rep 42:421–424
Monroe MC, Morzunov SP, Johnson AM et al (1999) Genetic diversity and distribution of Peromyscus-borne hantaviruses in North America. Emerg Infect Dis 5:75–86
Morzunov SP, Feldmann H, Spiropoulou CF et al (1995) A newly recognized virus associated with a fatal case of hantavirus pulmonary syndrome in Louisiana. J Virol 69:1980–1983
Morzunov SP, Rowe JE, Ksiazek TG et al (1998) Genetic analysis of the diversity and origin of hantaviruses in Peromyscus leucopus mice in North America. J Virol 72:57–64
Mustonen J, Partanen J, Kanerva K et al (1996) Genetic susceptibility to severe course of nephropathia epidemica caused by Puumala hantavirus. Kidney Int 49:217–221
Nemirov K, Vapalahti O, Lundkvist A et al (1999) Isolation and characterization of Dobrava hantavirus carried by the striped field mouse (Apodemus agrarius) in Estonia. J Gen Virol 80(Pt 2):371–379
Nemirov K, Henttonen H, Vaheri A et al (2002) Phylogenetic evidence for host switching in the evolution of hantaviruses carried by Apodemus mice. Virus Res 90:207–215
Nemirov K, Leirs H, Lundkvist A et al (2010) Puumala hantavirus and Myodes glareolus in northern Europe: no evidence of co-divergence between genetic lineages of virus and host. J Gen Virol 91:1262–1274
Nichol ST, Beaty BJ, Elliott RM et al (2005) Family bunyaviridae. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus taxonomy: VIIIth report of the international committee on taxonomy of viruses. Elsevier, London
Okumura M, Yoshimatsu K, Kumperasart S et al (2007) Development of serological assays for Thottapalayam virus, an insectivore-borne hantavirus. Clin Vaccine Immunol 14:173–181
Padula PJ, Colavecchia SB, Martinez VP et al (2000) Genetic diversity, distribution, and serological features of hantavirus infection in five countries in South America. J Clin Microbiol 38:3029–3035
Padula P, Della Valle MG et al (2002a) Andes virus and first case report of Bermejo virus causing fatal pulmonary syndrome. Emerg Infect Dis 8:437–439
Padula PJ, Sanchez AJ, Edelstein A et al (2002b) Complete nucleotide sequence of the M RNA segment of Andes virus and analysis of the variability of the termini of the virus S, M and L RNA segments. J Gen Virol 83:2117–2122
Pages M, Chaval Y, Herbreteau V et al (2010) Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. BMC Evol Biol 18:184
Pal E, Strle F, Avsic-Zupanc T (2005) Hemorrhagic fever with renal syndrome in the Pomurje region of Slovenia–an 18-year survey. Wien Klin Wochenschr 117:398–405
Pattamadilok S, Lee BH, Kumperasart S et al (2006) Geographical distribution of hantaviruses in Thailand and potential human health significance of Thailand virus. Am J Trop Med Hyg 75:994–1002
Pettersson L, Klingstrom J, Hardestam J et al (2008) Hantavirus RNA in saliva from patients with hemorrhagic fever with renal syndrome. Emerg Infect Dis 14:406–411
Plyusnin A, Morzunov SP (2001) Virus evolution and genetic diversity of hantaviruses and their rodent hosts. Curr Top Microbiol Immunol 256:47–75
Plyusnin A, Vapalahti O, Lankinen H et al (1994) Tula virus: a newly detected hantavirus carried by European common voles. J Virol 68:7833–7839
Plyusnin A, Cheng Y, Lehvaslaiho H et al (1996a) Quasispecies in wild-type tula hantavirus populations. J Virol 70:9060–9063
Plyusnin A, Vapalahti O, Lundkvist A et al (1996b) Newly recognised hantavirus in Siberian lemmings. Lancet 347:1835
Plyusnin A, Vaheri A, Lundkvist A (2003) Genetic interaction between Dobrava and Saaremaa hantaviruses: now or millions of years ago? J Virol 77:7156–7157, author reply 7157–7158
Plyusnin A, Vaheri A, Lundkvist A (2006) Saaremaa hantavirus should not be confused with its dangerous relative, Dobrava virus. J Clin Microbiol 44:1608–1609, author reply 1609–1611
Plyusnina A, Laakkonen J, Niemimaa J et al (2008a) Genetic analysis of hantaviruses carried by Myodes and Microtus rodents in Buryatia. Virol J 5:4

Plyusnina A, Laakkonen J, Niemimaa J et al (2008b) New genetic lineage of Tula hantavirus in Microtus arvalis obscurus in Eastern Kazakhstan. Open Virol J 2:32–36
Plyusnina A, Ferenczi E, Racz GR et al (2009a) Co-circulation of three pathogenic hantaviruses: Puumala, Dobrava, and Saaremaa in Hungary. J Med Virol 81:2045–2052
Plyusnina A, Ibrahim IN, Plyusnin A (2009b) A newly recognized hantavirus in the Asian house rat (Rattus tanezumi) in Indonesia. J Gen Virol 90:205–209
Raboni SM, Hoffmann FG, Oliveira RC et al (2009) Phylogenetic characterization of hantaviruses from wild rodents and hantavirus pulmonary syndrome cases in the state of Parana (southern Brazil). J Gen Virol 90:2166–2171
Ramsden C, Melo FL, Figueiredo LM et al (2008) High rates of molecular evolution in hantaviruses. Mol Biol Evol 25:1488–1492
Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol 26:143–153
Ravkov EV, Rollin PE, Ksiazek TG et al (1995) Genetic and serologic analysis of Black Creek Canal virus and its association with human disease and Sigmodon hispidus infection. Virology 210:482–489
Rawlings JA, Torrez-Martinez N, Neill SU et al (1996) Cocirculation of multiple hantaviruses in Texas, with characterization of the small (S) genome of a previously undescribed virus of cotton rats (Sigmodon hispidus). Am J Trop Med Hyg 55:672–679
Razzauti M, Plyusnina A, Henttonen H et al (2008) Accumulation of point mutations and reassortment of genomic RNA segments are involved in the microevolution of Puumala hantavirus in a bank vole (Myodes glareolus) population. J Gen Virol 89:1649–1660
Razzauti M, Plyusnina A, Sironen T et al (2009) Analysis of Puumala hantavirus in a bank vole population in northern Finland: evidence for co-circulation of two genetic lineages and frequent reassortment between strains. J Gen Virol 90:1923–1931
Reusken C, de Vries A, Adema J et al (2008) First genetic detection of Tula hantavirus in wild rodents in the Netherlands. J Infect 57:500–503
Reynes JM, Soares JL, Hue T et al (2003) Evidence of the presence of Seoul virus in Cambodia. Microbes Infect 5:769–773
Rivera P, Gonzalez Ittig R et al (2007) Molecular identification and phylogenetic relationships among the species of the genus Oligoryzomys (Rodentia, Cricetidae) present in Argentina, putative reservoirs of hantaviruses. Zool Scr 36:231–239
Rivers MN, Alexander JL, Rohde RE et al (2009) Hantavirus pulmonary syndrome in Texas: 1993–2006. South Med J 102:36–41
Rodier G, Soliman AK, Bouloumie J et al (1993) Presence of antibodies to hantavirus in rat and human populations of Djibouti. Trans R Soc Trop Med Hyg 87:160–161
Rollin PE, Ksiazek TG, Elliott LH et al (1995) Isolation of black creek canal virus, a new hantavirus from Sigmodon hispidus in Florida. J Med Virol 46:35–39
Salazar-Bravo J, Armien B, Suzan G et al (2004) Serosurvey of wild rodents for hantaviruses in Panama, 2000–2002. J Wildl Dis 40:103–109
Sanchez AJ, Abbott KD, Nichol ST (2001) Genetic identification and characterization of limestone canyon virus, a unique Peromyscus-borne hantavirus. Virology 286:345–353
Scharninghausen JJ, Pfeffer M, Meyer H et al (2002) Genetic evidence for tula virus in Microtus arvalis and Microtus agrestis populations in Croatia. Vector Borne Zoonotic Dis 2:19–27
Schlegel M, Klempa B, Auste B et al (2009) Dobrava-belgrade virus spillover infections, Germany. Emerg Infect Dis 15:2017–2020
Schmaljohn CS, Hooper JW (2001) Bunyaviridae: the viruses and their replication. In: Knipe DM, Howley PM (eds) Fields virology, 4th edn. Lippincott Williams & Wilkins, Philadelphia
Schmidt-Chanasit J, Essbauer S, Petraityte R et al (2010) Extensive host sharing of central European Tula virus. J Virol 84:459–474
Schultze D, Lundkvist A, Blauenstein U et al (2002) Tula virus infection associated with fever and exanthema after a wild rodent bite. Eur J Clin Microbiol Infect Dis 21:304–306
Schultze D, Fierz W, Matter HC et al (2007) Cross-sectional survey on hantavirus seroprevalence in Canton St. Gallen, Switzerland. Swiss Med Wkly 13:21–26

Schwarz AC, Ranft U, Piechotowski I et al (2009) Risk factors for human infection with Puumala virus, southwestern Germany. Emerg Infect Dis 15:1032–1039
Seijo A, Pini N, Levis S et al (2003) Study of hantavirus seoul in a human and rodent population from a marginal area in Buenos Aires City. Medicina (B Aires) 63:193–196
Sibold C, Meisel H, Kruger DH et al (1999) Recombination in Tula hantavirus evolution: analysis of genetic lineages from Slovakia. J Virol 73:667–675
Sibold C, Ulrich R, Labuda M et al (2001) Dobrava hantavirus causes hemorrhagic fever with renal syndrome in central Europe and is carried by two different Apodemus mice species. J Med Virol 63:158–167
Sironen T, Vaheri A, Plyusnin A (2005) Phylogenetic evidence for the distinction of Saaremaa and Dobrava hantaviruses. Virol J 2:90
Sironen T, Kallio ER, Vaheri A et al (2008) Quasispecies dynamics and fixation of a synonymous mutation in hantavirus transmission. J Gen Virol 89:1309–1313
Sjolander KB, Golovljova I, Vasilenko V et al (2002) Serological divergence of Dobrava and Saaremaa hantaviruses: evidence for two distinct serotypes. Epidemiol Infect 128:99–103
Sommer AI, Traavik T, Mehl R et al (1985) Reservoir animals for nephropathia epidemica in Norway: indications of a major role for the bank vole (C. glareolus) in comparison with the woodmouse (A. sylvaticus). J Hyg (Lond) 94:123–127
Song G (1999) Epidemiological progresses of hemorrhagic fever with renal syndrome in China. Chin Med J (Engl) 112:472–477
Song G, Hang CS, Liao HX et al (1984) Antigenic difference between viral strains causing classical and mild types of epidemic hemorrhagic fever with renal syndrome in China. J Infect Dis 150:889–894
Song JW, Baek LJ, Gajdusek DC et al (1994) Isolation of pathogenic hantavirus from white-footed mouse (Peromyscus leucopus). Lancet 344:1637
Song W, Torrez-Martinez N, Irwin W et al (1995) Isla Vista virus: a genetically novel hantavirus of the California vole Microtus californicus. J Gen Virol 76(Pt 12):3195–3199
Song JW, Baek LJ, Schmaljohn CS et al (2007a) Thottapalayam virus, a prototype shrewborne hantavirus. Emerg Infect Dis 13:980–985
Song JW, Gu SH, Bennett SN et al (2007b) Seewis virus, a genetically distinct hantavirus in the Eurasian common shrew (Sorex araneus). Virol J 4:114
Song JW, Kang HJ, Song KJ et al (2007c) Newfound hantavirus in Chinese mole shrew, Vietnam. Emerg Infect Dis 13:1784–1787
Song KJ, Baek LJ, Moon S et al (2007d) Muju virus, a novel hantavirus harboured by the arvicolid rodent Myodes regulus in Korea. J Gen Virol 88:3121–3129
Song JW, Kang HJ, Gu SH et al (2009) Characterization of Imjin virus, a newly isolated hantavirus from the Ussuri white-toothed shrew (Crocidura lasiura). J Virol 83:6184–6191
Tersago K, Verhagen R, Servais A et al (2009) Hantavirus disease (nephropathia epidemica) in Belgium: effects of tree seed production and climate. Epidemiol Infect 137:250–256
Tkachenko EA, Okulova NM, Iunicheva Iu V et al (2005) The epizootological and virological characteristics of a natural hantavirus infection focus in the subtropic zone of the Krasnodarsk territory. Vopr Virusol 50:14–19
Tomori O, Morikawa S, Matsuura Y et al (1986) Antibody to Japanese strain of hemorrhagic fever with renal syndrome (HFRS) virus in Nigerian sera. Trans R Soc Trop Med Hyg 80:1008–1009
Torrez-Martinez N, Hjelle B (1995) Enzootic of Bayou hantavirus in rice rats (Oryzomys palustris) in 1983. Lancet 346:780–781
Torrez-Martinez N, Bharadwaj M, Goade D et al (1998) Bayou virus-associated hantavirus pulmonary syndrome in Eastern Texas: identification of the rice rat, Oryzomys palustris, as reservoir host. Emerg Infect Dis 4:105–111
Vapalahti K, Paunio M, Brummer-Korvenkontio M et al (1999a) Puumala virus infections in Finland: increased occupational risk for farmers. Am J Epidemiol 149:1142–1151
Vapalahti O, Lundkvist A, Fedorov V et al (1999b) Isolation and characterization of a hantavirus from Lemmus sibiricus: evidence for host switch during hantavirus evolution. J Virol 73:5586–5592

Vapalahti O, Mustonen J, Lundkvist A et al (2003) Hantavirus infections in Europe. Lancet Infect Dis 3:653–661

Vapalahti K, Virtala AM, Vaheri A et al (2010) Case-control study on Puumala virus infection: smoking is a risk factor. Epidemiol Infect 138:576–584

Vincent MJ, Quiroz E, Gracia F et al (2000) Hantavirus pulmonary syndrome in Panama: identification of novel hantaviruses and their likely reservoirs. Virology 277:14–19

Vinh DC, Embil JM (2009) Hantavirus pulmonary syndrome: a concise clinical review. South Med J 102(6):620–625

Wang H, Yoshimatsu K, Ebihara H et al (2000) Genetic diversity of hantaviruses isolated in china and characterization of novel hantaviruses isolated from Niviventer confucianus and Rattus rattus. Virology 278:332–345

Wang ML, Lai JH, Zhu Y et al (2009) Genetic susceptibility to hemorrhagic fever with renal syndrome caused by hantaan virus in Chinese Han population. Int J Immunogenet 36:227–229

Xiao SY, Leduc JW, Chu YK et al (1994) Phylogenetic analyses of virus isolates in the genus Hantavirus, family Bunyaviridae. Virology 198:205–217

Yanagihara R, Daum CA, Lee PW et al (1987) Serological survey of Prospect Hill virus infection in indigenous wild rodents in the USA. Trans R Soc Trop Med Hyg 81:42–45

Yashina L, Abramove F, Gutorov V et al (2010) Seewis virus: phylogeography of a shrew-borne hantavirus in Siberia, Russia. Vector Borne Zoonotic Dis 10(6):585–591

Zeitz PS, Butler JC, Cheek JE et al (1995) A case-control study of hantavirus pulmonary syndrome during an outbreak in the southwestern United States. J Infect Dis 171:864–870

Zhang Y, Zhu J, Tao K et al (2002) Proliferation and location of hantaan virus in gamasid mites and chigger mites, a molecular biological study. Zhonghua Yi Xue Za Zhi 82:1415–1419

Zhang WY, Fang LQ, Jiang JF et al (2009) Predicting the risk of hantavirus infection in Beijing, People's Republic of China. Am J Trop Med Hyg 80:678–683

Zou Y, Hu J, Wang ZX et al (2008a) Molecular diversity and phylogeny of hantaan virus in Guizhou, China: evidence for Guizhou as a radiation center of the present hantaan virus. J Gen Virol 89:1987–1997

Zou Y, Hu J, Wang ZX et al (2008b) Genetic characterization of hantaviruses isolated from Guizhou, China: evidence for spillover and reassortment in nature. J Med Virol 80:1033–1041

Zou Y, Wang JB, Gaowa HS et al (2008c) Isolation and genetic characterization of hantaviruses carried by Microtus voles in China. J Med Virol 80:680–688

Zou Y, Xiao QY, Dong X et al (2008d) Genetic analysis of hantaviruses carried by reed voles Microtus fortis in China. Virus Res 137:122–128

Part V
New Integrations

Chapter 10
Nasal carriage of Methicillin-Resistant Coagulase-Negative Staphylococci: A Reservoir of *mecA* Gene for *Staphylococcus aureus*

Raymond Ruimy, François Barbier, David Lebeaux, Etienne Ruppé, and Antoine Andremont

Abstract The efficacy of penicillin in *Staphylococcus aureus* infection has justified its widespread use. The counterpart has been the rapid emergence of penicillin resistant strains. Currently, almost 90% of *S. aureus* strains are penicillin resistant by the acquisition of a plasmid-encoded penicillinase. The use of a new semi-synthetic β-lactam, the methicillin has allowed treating these resistant strains. However, as for the penicillin, methicillin resistant *S. aureus* (MRSA) have rapidly been selected but remained mainly confined to hospitals. Recently, MRSA have independently emerged in the community. The acquisition of *mecA* gene carried on a mobile genetic element called Staphylococcal Cassette Chromosome *mec* (SCC*mec*) confers methicillin resistance. This gene encodes for a penicillin-binding protein which has a low affinity not only for most semi-synthetic penicillin (such as methicillin) but also for the entire β-lactam class. The origin of this gene and cassette are unknown. However, SCC*mec* is widespread in coagulase negative *staphylococcus* (CoNS). Since methicillin susceptible *S. aureus* and CoNS share the same ecological niches

R. Ruimy (✉) • D. Lebeaux • E. Ruppé • A. Andremont
Bacteriology Unit and National Reference Center for Emergence of Resistance in Commensal Flora, Bichat-Claude Bernard Hospital (Assistance Publique-Hôpitaux de Paris), Paris, France

EA 3964, Xavier Bichat Medical School (Denis Diderot-Paris 7 University), Paris, France
e-mail: raymond.ruimy@bch.aphp.fr; david.lebeaux@yahoo.fr; etienne.ruppe@bch.aphp.fr; antoine.andremont@bch.aphp.fr

F. Barbier
EA 3964, Xavier Bichat Medical School (Denis Diderot-Paris 7 University), Paris, France

Medical Intensive Care Unit, Bichat-Claude Bernard Hospital (Assistance Publique-Hôpitaux de Paris), Paris, France
e-mail: francois.barbier@bch.aphp.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*, 219
DOI 10.1007/978-94-007-2114-2_10,

in humans such as the anterior nares, transfer could occur from MR-CoNS to methicillin susceptible *S. aureus* but its mechanism remains unknown. In this review, we will describe what argue in favor of this hypothesis, focusing on what is currently known on the structural genetic organization of SCC*mec* found in MR-CoNS and MRSA.

10.1 Introduction

Staphylococcus aureus is recognized as a major human pathogen causing a wide range of infections but is also a commensal species colonizing the skin and mucous membranes of humans. *S. aureus* infections are ranging from skin infections usually benign to bloodstream infections which are frequently life threatening (Lowy 1998). They can occur both in the community and hospitals. *S. aureus* is also able to asymptomatically colonise healthy individuals. Among the multiple body sites where it can be detected, the anterior nares are the most frequent constituting an important ecological reservoir of this organism in humans. At any given point in time and any country, 15–40% of individuals are asymptomatic nasal carriers of *S. aureus* (Wertheim et al. 2005). Nasal carriers have nevertheless at higher risk of infection than non-carriers (von Eiff et al. 2001; Wertheim et al. 2004). *S. aureus* nasal detection has become the gold standard to determine whether a patient is a carrier or not. Also, it has been demonstrated that nasal eradication of the anterior nares, which can be obtained using topical antibiotics, significantly reduce the risk of infection in carriers (Bode et al. 2010). Unfortunately, however, this is a practice associated with emergence of resistant strains (Perl et al. 2002).

Prior to the "antibiotic era", *S. aureus* was naturally susceptible to all antibiotics. It represents the first typical example of the impact of widespread antibiotic use on the selection of resistant bacteria. The history of the emergence of antibiotic resistance in *S. aureus* has recently be analysed as a series of waves (Chambers and Deleo 2009). In 1940, the use of penicillin has shown to be highly effective against *S. aureus* infection. The penicillin has therefore been largely used. However, 2 years later, the first isolate resistant to this antibiotic has been described first in hospitals, next in the community (Barber 1947; Barber and Rozwadowska-Dowzenko 1948; Kirby 1944). The prevalence of penicillin-resistant strains has rapidly increased within few years to reach 90% of all strains of the species. The mechanism of penicillin resistance is due to the acquisition of a plasmid-encoded penicillinase hydrolysing the β-lactam ring of penicillin. The origin of this gene is unknown. In 1960, a new semi-synthetic antibiotic, methicillin, has been introduced to treat infections caused by penicillin-resistant *S. aureus* strains. However, 1 year later, the first strain of methicillin resistant *S. aureus* (MRSA) has been observed (Barber 1961; Jevons 1961). During 50 years, the prevalence of MRSA strains has increased slowly and then rapidly from late 1980 in most hospitals worldwide and has reached in some of them 30–60% of all hospitals *S. aureus*. However, several countries in northern Europe have a low prevalence of MRSA in hospital. Recently, several countries such

as France have reported downward trends (Jarlier et al. 2010). MRSA strains remained mainly confined to hospitals and other institutional health care settings causing frequent infections in persons who have a history of recent hospitalization or who had a close contact with an individual recently hospitalized, or working in hospitals. They were called healthcare-associated MRSA (HA-MRSA) strains. The vast majority of HA-MRSA has as characteristics to be also resistant to non-β-lactam antibiotics class such as lincosamides, macrolides, aminoglycosides, fluoroquinolones, or combinations of theses antibiotic. They belong for most of them to ten major pandemic clones named Archaic, Berlin, Iberian, Irish-1, Brazilian, Hungarian, New York/Japan, Southern Germany, several UK EMRSA, and Pediatric clones (see for review Deurenberg and Stobberingh (2008)). Since 1990, new MRSA clones have emerged in community causing infections in individual without exposure to healthcare and have been called community associated MRSA (CA-MRSA) (Anonymous 1999; Coombs et al. 2004; Herold et al. 1998; O'Brien et al. 2004; Udo et al. 1993). The causes of their emergence remains unknown, especially in populations exposed to low selective antibiotic pressure comparatively to what is observed in the hospital setting. CA-MRSA differ phenotypically and genotypically from HA-MRSA (Deurenberg and Stobberingh 2008). For example, the majority of CA-MRSA isolates are more susceptible to non-β-lactam antibiotics than HA-MRSA. The genetic diversity of CA-MRSA has been found to be greater than that of HA-MRSA suggesting that more *S. aureus* susceptible lineage became CA-MRSA (Tristan et al. 2007; Vandenesch et al. 2003). They have rapidly been reported in several countries including those known to have a low prevalence of MRSA, such as countries from northern Europe (Bartels et al. 2007). In the United States, CA-MRSA have rapidly diffused in the community and become the major cause of skin and soft tissue infections (Moran et al. 2006).

In contrast to *S. aureus*, the others members from the genus *Staphylococcus* cause rarely infections in humans except for patients who carry foreign body materials and/or who are immunocompromized. Among the 48 species contained in the genus, only a few are isolated from humans. Since the majority of them are unable to coagulate rabbit plasma, they are grouped together under the generic name of coagulase-negative staphylococci (CoNS). In humans, three species are the most frequent: *Staphylococcus epidermidis, Staphylococcus haemolyticus*, and *Staphylococcus hominis*. CoNS differ also from *S. aureus* by the high prevalence of carriage in commensal flora but both colonise the same body site: the anterior nares (Becker et al. 2006; Costa et al. 2004). Lastly, CoNS are usually much more resistant to antibiotics than *S. aureus*.

The mechanism of methicillin resistance has been elucidated more than 20 years after the first description of a MRSA strain (Hartman and Tomasz 1984; Ubukata et al. 1985). It is due to acquisition of *mecA* gene encoding for a penicillin-binding protein (PBP) 2a or PBP 2′, a transpeptidase which has a low affinity not only for most semi-synthetic penicillin (such as methicillin) but also for the entire β-lactam class of antibiotics. Indeed, these are ineffective for treatment of MRSA infections (Matsuhashi et al. 1986). *mecA* gene is carried on a mobile genetic element called Staphylococcal Cassette Chromosome *mec* (SCC*mec*) (Ito et al. 1999; Katayama et al. 2000)

which was found inserted into the chromosome of susceptible strains from several *S. aureus* and CoNS lineages. Soon after the first description of SCC*mec*, several different genetic structures of SCC*mec* were found in different *Staphylococcus* species; SCC*mec* diversity being higher in methicillin resistant CoNS (MR-CoNS) species than in MRSA. Recent reports have documented that nasal carriage prevalence of MR-CoNS was higher than that of MRSA in the community, as well as in the hospital setting. MR-CoNS could then act as a potential *mecA* gene reservoir for methicillin susceptible *S. aureus* (MSSA). Since CoNS and MSSA share the same ecological niches in humans such as the anterior nares, SCC*mec* transfer could occur from MR-CoNS to MSSA but its mechanism remains unknown. In this review, we will describe what argues in favor of this hypothesis, focusing on what is currently known on the structural genetic organization of SCC*mec* found in MR-CoNS and MRSA.

10.2 SCC*mec* Element: Structural Genetic Organization and Classification

SCC*mec* element is a DNA fragment which was initially called "additional DNA" or "mecDNA" since it carries *mecA* and has no allelic equivalent in MSSA strains (Ito et al. 1999; Katayama et al. 2000). Their sizes range from 21 to 67 kb. It has been found localized on the chromosome of *S. aureus* between the genes encoding protein A (*spa*) and a protein involved in the biosynthesis of purine (*purA*), near the chromosomal origin of replication. It is integrated at a specific site, designated *attBscc*, located at the 3′ extremity of *orfX*, a gene of unknown function (Ito et al. 1999). SCC*mec* is flanked by direct repeat (DR) sequences containing integration site sequences. The recombination events (insertion and excision) of SCC*mec* in *S. aureus* chromosome are driven by specific recombinase genes designated *ccr* (for cassettes chromosom recombinases) and carried by SCC*mec* (Katayama et al. 2000). The Ccr recombinases are considered as belonging to the invertase/resolvase family since their NH2-terminal is highly homologous to those of members of this family (Abdel-Meguid et al. 1984). They clearly differ from the integrase family of site-specific recombinases, such as those of bacteriophage lambda and conjugative transposons. Consequently, SCC*mec* has been described as a novel genetic mobile element carrying *mecA*. Although SCC*mec* is widely distributed in several members of the *Staphyloccus* genus, the mechanism of SCC*mec* transfer from cell to cell remains unknown to date.

SCC*mec* elements are divided in three so-called J regions and two genes complexes, namely the *mec* and *ccr* gene complexes (Fig. 10.1). Recently, it has been proposed that the term J region should refer to "joining region" but not to "junkyard region", as previously designed (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements 2009). The three J regions are located for J1, between the right chromosomal junction and the *ccr* gene complex, for J2 between *ccr* gene complex and the *mec* gene complex, and for J3 between the *mec* complex and the left chromosomal junction (*orfX*). Although these regions may

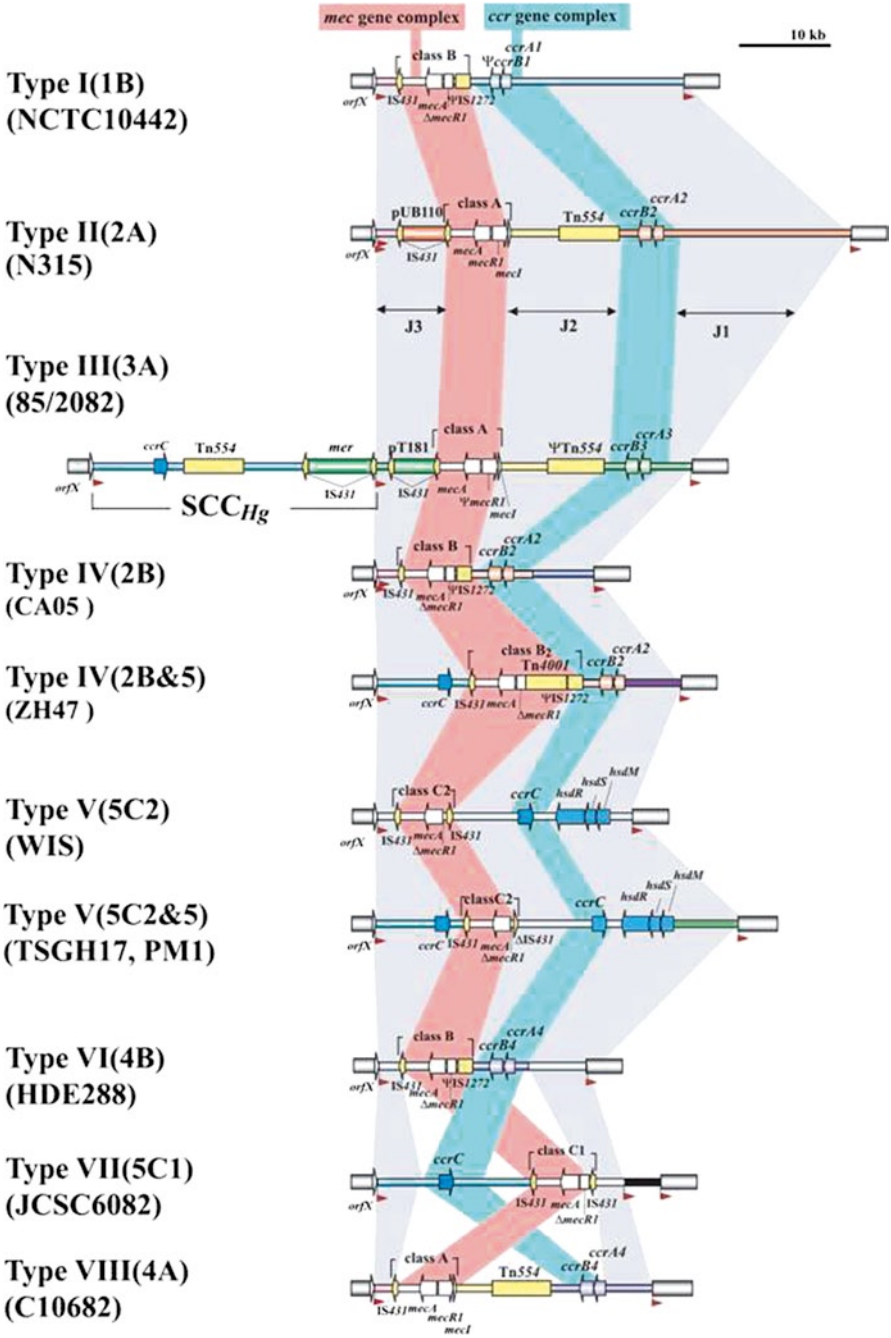

Fig. 10.1 Structural comparison of SCC*mec* elements which are composed of the *ccr* gene complex (*blue*) and the *mec* gene complex (*peach*). Other areas (*light gray*) of SCC*mec* are nonessential and are divided into three regions, J1–J3. The *ccr* gene complex consists of *ccr* genes that are responsible for the mobility of SCC*mec* and surrounding ORFs. The *mec* gene complex is responsible for methicillin-cephem resistance. *Red arrowheads* indicate the IS of SCC that comprise DR sequences. (From International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (2009). With permission)

carry genes encoding additional antimicrobial resistance, they are not considered as essential components of the cassette although they play a role in the hierarchical system classification of SCC*mec*. Recently, the classification of SCC*mec* elements has been revised by an International Working group on the classification of staphylococcal cassette chromosome elements (IWG-SCC) (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements 2009). The definition of a SCC*mec* type is based on the combination of a *ccr* gene complex represented by the *ccr* gene allotype and of the class of the *mec* gene complex. A subtype of SCC*mec* is defined by the variations in the J regions.

The *ccr* gene complex includes *ccr* genes which belong to three distinct phylogenetic groups designed *ccr*A, *ccr*B, and *ccr*C sharing less than 50% of DNA similarity with each other. To date, four allotypes of *ccr*A and *ccr*B and only one of *ccr*C have been found in *S. aureus*. They are designed as *ccr*A1 to A4, *ccr*B1 to B4 and *ccr*C1. Their association defined the 5 types of the *ccr* gene complexes found in *S. aureus*, as follows: type 1 carrying *ccr*A1B1, type 2 carrying ccrA2B2, type 3 carrying ccrA3B3, type 4 carrying *ccr*A4B4 and type 5 carrying ccrC1. An additional allotype designed as *ccr*A5 have been identified in *Staphylococcus pseudointermedius* (Descloux et al. 2008) and still others, designated *ccr*B6 and *ccr*B7, in *Staphylococcus saprophyticus* (Higashide et al. 2008). The IWG-SCC has recently proposed to recognize a novel *ccr* and a novel allotype on the basis of the percentage of DNA sequence similarity with the previously described *ccr* (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements 2009). The cut off limit of DNA sequence similarity is below 50% for a novel *ccr* and between 50% and 85% for a novel *ccr* allotypes.

The *mec* gene complex is composed of *mec*A, its regulatory genes and associated insertion sequences. The regulation of *mec*A is driven by the repressor MecI encoded by *mecI* gene and the trans-membrane β-lactam-sensing signal transducer MecR1, encoded by *mecR*. In the absence of a β-lactam antibiotic, MecI represses both the transcription of *mec*A and *mec*R1-*mec*I. In the presence of a β-lactam antibiotic, MecR1 is auto-catalytically cleaved, and the metalloprotease domain, located in the cytoplasmic part of MecR1, becomes active and cleaves MecI which is bound to the *mecA* operator region (Chambers 1997). As a consequence, the *mec*A gene can be transcripted and, subsequently, production of PBP2a occurs. Four classes of *mec* genes complex have been identified and designed by A to D. The prototypical class A *mec* gene complex contains *mecA*, the complete *mecR1*, *mecI* genes upstream of *mecA*, and insertion sequence IS*431* downstream of *mecA* (Katayama et al. 2001). The class B *mec* gene complex contains *mecA*, a truncated *mecR1* by the insertion of IS*1272* upstream of *mecA*, and IS*431* downstream of *mecA* (Heusser et al. 2007). The class C *mec* gene complex contains *mecA*, a *mecR1* gene truncated by the insertion of IS*431* upstream of *mecA*, and IS*431* downstream of *mecA* (Kobayashi et al. 2001). Two classes C *mec* gene complexes, class C1 and C2, have been separated upon the orientation of IS*431* upstream of *mecA*. They have likely evolved independently. The IS*431* upstream of *mecA* has the same orientation than when downstream of *mecA* in the class C1 complexes while it reversed in class C2. The class D *mec* gene complex contained *mecA* and Δ*mecR1* (no insertion sequence downstream of Δ*mecR1* has been found) (Suzuki et al. 1993).

Since additional IS, such as IS*431* and IS*1182*, have been found inserted within class A *mec* gene complexes, several variants of class A have been described and designed by a numerical string following the class A: class A_2, A_3 and A_4. The class A_2 carries IS*431* downstream of *mecI* and has been found in *S. haemolyticus*. The classes A_3 and A_4 have been identified in MRSA strains carrying IS*1182* in and around the *mecI* gene, respectively. A variant of class B, designed as class B_2 *mec* gene has also been described due to the presence Tn*4001* inserted upstream of *mecA*.

To date, the methods for typing SCC*mec* are based on multiplex PCR, such as proposed by Kondo et al. (Kondo et al. 2007). This PCR-scheme requires using five multiplex PCR. This approach is not easy to perform routinely but is the most appropriate to type SCC*mec*. However, determining the entire nucleotide sequence is required to describe a novel SCC*mec*.

In all, SCC*mec* types have firstly been designed by roman numbers only. Although this nomenclature is non informative on the *ccr* allotype or class of *mec* gene complexes, the IWG-SCC has proposed to retain it because it has widely been used to type SCC*mec* from MRSA clones (International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements 2009). However, in order to be more informative, each SCC*mec* type should also be designated by an additional couple of an Arabic number and a capital letter in parentheses defining the *ccr* and *mec* gene complexes harboured by the SSC*mec* type, Type I for 1B for example (Fig. 10.1). To design the SCC*mec* subtype, three methods have been used to define the differences of J1, J2, and J3 regions but the IWG-SCC has recently proposed to use a system developed by Stephens et al. (Stephens et al. 2007), as provided on a dedicated website: http://www.sccmec.org (IInternational Working Group on the Classification of Staphylococcal Cassette Chromosome Elements 2009).

10.3 SCC*mec* Element in *S. aureus*

It has experimentally been shown that, if SCC*mec* is carried an a multicopy plasmid, it transfers into the chromosom in a site and orientation-specific manner, driven by the recombinases carried by SCC*mec* (Katayama et al. 2000). However, the precise mechanism by which SCC*mec* is transferred in Nature to *S. aureus*, the conditions promoting this transfer and the nature of the original SCC*mec* reservoir are so far unknown. However, several researchers have estimated the number of times that SCC*mec* has been acquired by *S. aureus* on the basis of evolutionary models (Nubel et al. 2008; Robinson and Enright 2003). The construction of such models has been made possible since the strains have unambiguously been identified by their partial sequences of seven housekeeping genes (Enright et al. 2000). This molecular typing method is called multi-locus sequence typing (MLST) (Enright et al. 2000). In the early 2000s years, using multilocus sequence typing (MLST) and the eBURST algorithm on a international collection of strains, 11 major MRSA clones have been identified within five phylogenetically distinct lineages which led to the hypothesis

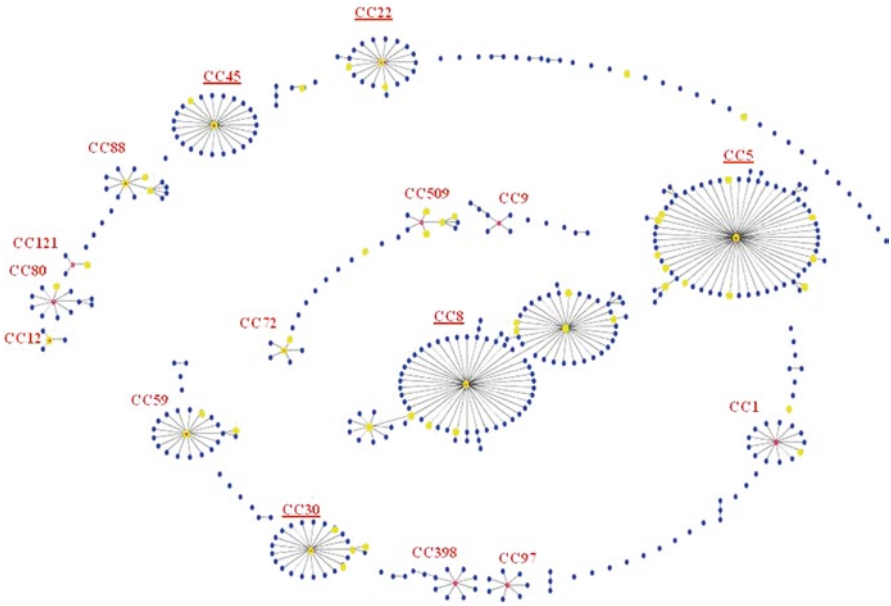

Fig. 10.2 Analysis of the entire methicillin resistant *Staphylococcus aureus* multi-locus sequence typing database (Accessible at www.saureus.mlst.net) using e-Burst algorithm (Feil et al. 2004). For clarity, ST labels have been removed. STs published before and after 2002 are shown as *yellow* and *blue circle*. The predicted ST founder of each clonal complex have shown as red circles. Only clonal complexes designated as CC are labeled. The CC underlined are those described in 2002

that MRSA emerged on multiple occasions within a given phylogenetic lineage (Enright et al. 2002; Robinson and Enright 2003). The five phylogenetic lineages are recognized by eBURST as clonal complexes (CCs). A given clonal complex includes all STs which have, at least, five to seven loci identical with a predicted ancestral allelic profile (Feil et al. 2004). CC5, CC8, CC22, CC30, and CC45 constitute the five major MRSA CCs (Fig. 10.2). To estimate more accurately *mecA* acquisition frequency by *S. aureus*, a higher-resolution multilocus-locus sequence typing has been developed, targeting the seven originally used house keeping genes and eight new sequences (seven from genes encoding surface proteins known to be more polymorphic and repeat sequences of the *spa* gene) (Robinson and Enright 2003). For each linage, an evolutionary model has been constructed on the basis of these data to estimate the frequency of MRSA emergence. The authors of this work have proposed that MRSA has emerged at least 20 times in *S. aureus* (Robinson and Enright 2003). More recently, a new evolutionary scenario has been proposed estimating that the acquisition frequency of SCC*mec* by MSSA was approximately 20 times higher than previously determined (Nubel et al. 2008). This scenario was deduced from analysis of a single ST, ST5, by single nucleotides polymorphisms obtained within 108 genes loci spanning 46 kb from each isolate within a collection of 135 isolates. Isolates were collected from 22 countries in six continents (Nubel et al. 2008). The results showed that acquisition of SCC*mec* by MSSA ST5

occurred at least 23 times and that the potential progeny of such strains were usually distributed locally rather than globally. The perception that only few MRSA clones have spread globally and that the acquisition of SCC*mec* by *S. aureus* was a rare event must thus be challenged by these results on ST5, especially if they can be extended to other STs. It is now believed that the acquisition frequency of SCC*mec* by *S. aureus* is higher than previously thought (Robinson and Enright 2003). Furthermore, the number of STs, and to a lesser degree of CCs, has considerably increased since the first scenario. In comparison to 2002 (Enright et al. 2002), the number of ST and CCs increased from 54 and 5 to 438 and 16, respectively (Fig. 10.2). This increase in description of new STs reinforces the idea that the number of acquisition is still underestimated. However, each in construction might not be epidemiologically successful.

To date, eight main SCC*mec* allotype designated SCC*mec* I to VIII have been described in MRSA clones (Berglund and Soderquist 2008; Ito et al. 2001; Ito et al. 1999; Ito et al. 2004; Ma et al. 2002; Oliveira et al. 2006; Zhang et al. 2009) (Fig. 10.1). MRSA clones are best characterised by their ST types and SCC*mec* elements. The first HA-MRSA described harbour SCC*mec* I to III which were considered to be the longest SCC*mec* (31–67 kb) whereas the first CA-MRSA described carried SCC*mec* IV which is the smallest SCC*mec* (21–24 kb). Since SCC*mec* IV has been rarely found in MRSA before 1990, it has been suggested that this small SCC*mec* has recently been acquired by *S. aureus*, probably from another staphylococci species, such as *S. epidermidis*, because it is the more prevalent MR-CoNS species (Naimi et al. 2001; Okuma et al. 2002; Oliveira et al. 2002).

10.4 Origin of *mecA* and SCC*mec* Element

Since the first description of MRSA, finding the origin of *mecA* gene and of SCC*mec* has been a challenge for microbiologists. It is not resolved yet but some hypotheses have been proposed. The exogenous origin of SCC*mec* has been suggested because of its GC% composition and its codon usage which are different from those of genomic DNA, and of the presence of direct and inverted repeat which flanked SCC*mec* (Hiramatsu et al. 2001). The insertion of SCC*mec* at a specific site in *orfX* and the presence of a recombinase carried by SCC*mec* also concur to consider that SCC*mec* is a mobile genetic element originating from an another species. The *mecA* gene carried by SCC*mec* from MRSA and from MR-CoNS showed 99.9% DNA sequence similarity between strains, suggesting a unique source (Wisplinghoff et al. 2003) (Fig. 10.3). The screening of several CoNS by hybridization using a *mecA* DNA probe has revealed that *Staphylococcus sciuri* could carry a homologue of MRSA *mecA* (Wu et al. 1996). It appeared as a good donor candidate since it is considered as one of the most abundant staphylococcal species on our planet (Kloos et al. 1997). Two additional arguments have supported this hypothesis (Couto et al. 1996; Couto et al. 2000; Wu et al. 1996). The *S. sciuri mecA* showed 79.5% DNA sequence similarity to the *mecA* gene of MRSA revealing it is closely

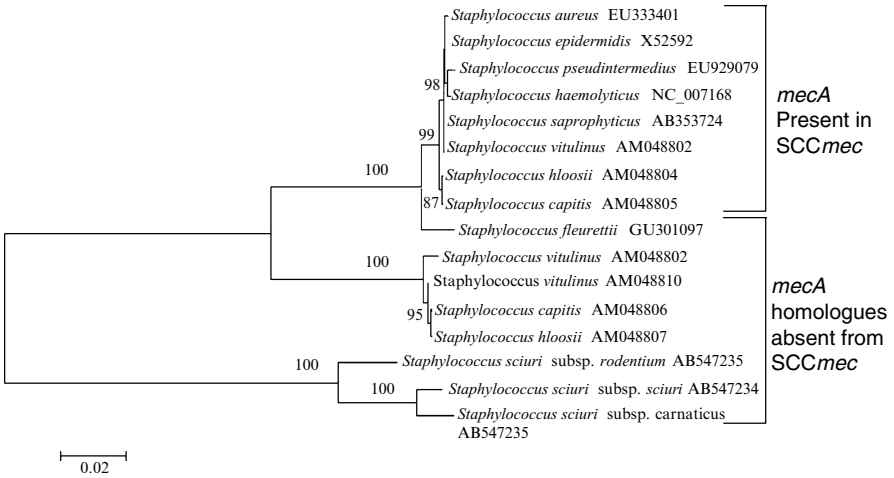

Fig. 10.3 Neighbor-joining tree based on the comparison of 2007-bp *mecA* gene carried by SCC*mec* element from *S. aureus* and others staphylococci, and of its *mecA* gene homologue carried outside SCC*mec* from *Staphylococcus* spp.

related (Fig. 10.3). An even higher level of identity (91.2%) has been found between the deduced amino acid sequences from two *mecA* when their transpeptidases domains were compared. In addition, *S. sciuri mecA* homologue has been detected in all *S. sciuri* isolates from a wide variety of ecological sources, suggesting that it was native in this species. Although *S. sciuri mecA* gene was present in all strains of the species, strains were fully susceptible to beta-lactam antibiotics, except for the isolates selected in vitro or from rare human clinical sources which had a mutation or an *IS* element inserted in the promoter region of the *mecA* gene (Couto et al. 2003). Recently, a third variant type of *mecA* gene has been found in three CoNS species: *Staphylococcus kloosii*, *Staphylococcus vitulinus*, and *Staphylocccus capitis* (Schnellmann et al. 2006). All were isolated from horse skin. This new *mecA* gene, which displayed 91% DNA identity to the *mecA* MRSA gene, is more closely related to *mecA* MRSA gene than that carried by *S. sciuri* (Fig. 10.2). However, it was not found in all isolates from these three species. More recently, a *mecA* gene, highly homologous to that of MRSA (99.8% nucleotide identity) has been found in several strains of *Staphylococcus fleurettii* (Tsubakishita et al. 2010). This species phylogenetically grouped by 16S rRNA analysis with three other species i.e. *S. sciuri*, *S. vitullinus* and *Staphylococcus lentus*. This group has been called *S. sciuri* species group. To date, only strains of the species *S. lentus* did not carry *mecA* homologue in their chromosome. Since all strains of this *S. sciuri* group are more frequently isolated from a variety of animals and food products of animal origin than from humans, CoNS species carrying *mecA* homologue and colonizing animal may play an important role as the source of *mecA* for *S. aureus*. In all, several lines of findings suggest that the evolutionary precursor of *mecA* MRSA gene could be carried by the chromosome of CoNS species not associated with SCC*mec*.

The source of the others constituents of the *mecA* gene complex or SCC*mec* remained elusive since they were absent surrounding the chromosomal region of the *mecA* precursor from *S. sciuri* and *S. vitullinus* (Tsubakishita et al. 2010). More recently, the sequencing of *mecA* region from *S. fleurettii* strains has shown the presence of a stretch of a chromosomal sequence (~12 kb long) that had 99% nucleotide identity with the corresponding region of the class A *mec* gene complex carried by SCC*mec* type II, III, and VIII carried by MRSA. This suggests that the formation of these SCC*mec* types required not only *S. fleurettii* but also a donor strain of the others component of SCC*mec*, such as recombinases and *IS* responsible for the mobilisation and insertion of *SCCmec*.

It is interesting to note that all the *S. fleurettii* strains carrying *mecA* are resistant to oxacillin while those of *S. sciuri*, which all carry a *mecA* precursor, are mostly susceptible. All these findings suggest that *S. fleurettii* could play a role as the source of the *mecA* donor and for class A *mec* gene complexes. The acquisition of *mecA* by some strains of *S. fleurettii* which is a commensal host of animals may due to antibiotic selective pressure resulting from antibiotic exposure of animals. To find the animal *Staphylococcus* species in which SCC*mec* originated as a mobile genetic element is a new challenge. The hypothesis that *mecA* and SCC*mec* originated from commensal CoNS strains of animals seems attractive. However, these species are rarely isolated from humans. The transfer of SCC*mec* from CoNS to *S. aureus* in animals followed by the transmission of MRSA from animals to humans is a possibility. However, MRSA found in animals are rarely transmitted to humans and their types are mainly animal specific (Cuny et al. 2010). This has been very well demonstrated in the case of *S. aureus* MRSA ST398 (Armand-Lefevre et al. 2005; Huijsdens et al. 2006). Indeed, MRSA ST398 has been isolated from pigs but also from pig farmers (Armand-Lefevre et al. 2005) and their human contacts (Huijsdens et al. 2006), suggesting that some of animals MRSA could have a limited host specificity and be able to colonise human. However, these strains do not disseminate in humans far away from pigs settings and only few and limited human epidemics have been described.

10.5 Reservoir of SCC*mec* Elements in CoNS Species Colonizing Humans

S. epidermidis is the most prevalent CoNS species among the staphylococcal flora colonizing the normal human skin and mucosa including the nasal microbiota, followed by *S. haemolyticus* and *S. hominis* (Kloos 1997; Kloos and Musselwhite 1975). The nasal carriage of MR-CoNS has been studied in different groups of subjects according to their level of exposition to the health care system. In hospitalized patients, the prevalence of MR-CoNS nasal carriage is currently >40% and can reach 80% (Archer and Armstrong 1983; Krediet et al. 2001; Lee et al. 2000). Antibiotic pressure and cross contamination are the two main factors which increase the carriage of MR-CoNS in hospitalized patients (Terpstra et al. 1999). In outpatients, several reports have also emphasized that the prevalence of MR-CoNS nasal carriage

was high (Barbier et al. 2010; Jamaluddin et al. 2008; Ruppe et al. 2009; Silva et al. 2001). We found that in outpatients living in countries with contrasting environments (Algeria, Mali, Moldova and Cambodia) the carriage rate ranged from 11% to 31% with significant variations between the four countries (Ruppe et al. 2009). As for MRSA nasal carriage (Lucet et al. 2003), previous exposition to the health care system (defined as hospitalization in the previous year, or long-term hemodialysis, or nursing at home, or living in a rest home), was associated with a higher prevalence of MR-CoNS nasal carriage (Barbier et al. 2010). Recently, it has been shown that in outpatients previously exposed to the health care system, the carriage rate of MR-CoNS reached 25.9% while it decreased to 16.5% in non-exposed patients (Barbier et al. 2010). Others have found that in subjects living in the community, the rate of carriage ranged from 25% to 30.35% (Jamaluddin et al. 2008; Silva et al. 2001).

The biodiversity of SCC*mec* elements carried by MR-CoNS strains isolated from human nasal microbiota has been studied in different populations (Barbier et al. 2010; Hanssen and Sollid 2007; Jamaluddin et al. 2008; Miragaia et al. 2008; Miragaia et al. 2005; Ruppe et al. 2009). It has been found to be much more diverse than these from MRSA isolates. Four types of SCC*mec* have been found in MR-CoNS: (*i*) SCC*mec* which can be assigned to the SCC*mec* described for *S. aureus*, (*ii*) SCC*mec* with new associations between *mec* and *ccr* complexes, (*iii*) SCC*mec* with combinations of several *mec* or *ccr* complexes, and (*iv*) untypeable SCC*mec*. SCC*mec* which have been best characterized were those carried by MR-*S. epidermidis* because this species is the most prevalent among nasal MR-CoNS strains, as well as among infection caused by MR-CoNS. Among this species, SCC*mec* type IV has been found in all studies as the more prevalent, with rate a ranging from 27% to 41% (Barbier et al. 2010; Hanssen and Sollid 2007; Jamaluddin et al. 2008; Miragaia et al. 2008; Miragaia et al. 2005; Ruppe et al. 2009). Their nucleotide sequences had 98–99% homology with SCC*mec* type IV carried by MRSA (Wisplinghoff et al. 2003). More recently, complete sequences of SCC*mec* type IVa from 3 MR-*S. epidermidis* strains have confirmed their high homology (>99%) with those available for CA-MRSA, including the major clones USA300 and USA400 (Barbier et al. 2010). These results support the hypothesis that SCC*mec* has been exchanged between S. *epidermidis* and *S. aureus*. Concerning the others SCC*mec* types, SCC*mec* type I prevalence was low ranging from 0% to 4% whereas the prevalence of SCC*mec* type II, III and V were much more variable, ranging from 0% to 29.6%, 0% to 29.6% and 0% to 12%, respectively. Last, untypeable SCC*mec* were frequent in MR-*S. epidermidis* ranging from 18% to 51%. These untypeable SCC*mec* included some with an association between one *mec* complex and multiple *ccr*, suggesting that several homologous recombinations between distinct SCC*mec* elements has occurred, as previously suggested for SCC*mec* VIII in MRSA strain C10682 (Zhang et al. 2009). It has been suggested that these sequential insertions of foreign genetic elements in *orfX* contribute to the frequent generation of new SCC*mec* elements resulting in the genomic diversity of *S. epidermidis* (Gill et al. 2005; Zhang et al. 2003). These data constitute one line of evidence supporting the role of MR-*S. epidermidis* as a reservoir of SCC*mec* for *S. aureus*. SCC*mec* in others MR-CoNS has been studied in *S. haemolyticus,*

S. hominis, S. saprophyticus, S. cohnii, S. pseudointermedius and *S. pettenkoferi* (Barbier et al. 2010; Berglund and Soderquist 2008; Descloux et al. 2008; Ruppe et al. 2009; Soderquist and Berglund 2009). *S. haemolyticus, S. hominis, S. saprophyticus*, and *S pettenkoferi*, all carried some typeable SCC*mec*. SCC*mec* type V predominated in *S. haemolyticus* from humans living in various countries (Barbier et al. 2010; Berglund and Soderquist 2008; Ruppe et al. 2009). Partial sequencing of their *mec* complexes has revealed over 99% nucleotide identity with those of SCC*mec* type V described in MRSA, suggesting an SCC*mec* exchange between *S. haemolyticus* and *S. aureus* (Berglund and Soderquist 2008). The other SCC*mec* of *S. haemolyticus* have been found untypeable and highly diverse as in *S. epidermidis*. Recently, it has been found that some isolates from *S. saprophyticus, S. hominis* and *S pettenkoferi* carried SCC*mec* type III, SCC*mec* type VIII and SCC*mec* type IV, respectively (Barbier et al. 2010; Soderquist and Berglund 2009).

In all, the high prevalence of MR-CoNS nasal carriage in the community, just as in hospitalized patients, and the high diversity of SCC*mec* carried by MR-CoNS support the hypothesis that MR-CoNS constitutes a pool of SCC*mec* which can be transfer to MSSA, the transfer being probably mediated by the expression of *ccr* genes.

10.6 SCC*mec* Transfer to MSSA

Although, the precise mechanisms of SCC*mec* transfer between MR-CoNS to MSSA have not yet been elucidated, several lines of evidence support the hypothesis. Several clinical reports involving health care-associated strains suggest that SCC*mec* transfer from MR-CoNS to methicillin susceptible *S. aureus* may occur *in vivo* (Berglund and Soderquist 2008; Ibrahem et al. 2009; Jamaluddin et al. 2008; Wielders et al. 2001; Wisplinghoff et al. 2003; Zhang et al. 2009). The first clinical observation of the transfer of SCC*mec* from CoNS to *S. aureus* has been reported in an infant treated with beta-lactam antibiotics (amoxicillin and clavulanic acid followed by flucloxacillin) (Wielders et al. 2001). The transfer has been hypothesized to have occurred *in vivo* because closely related MSSA and MRSA, and a potential SCC*mec* donor *S. epidermidis* strain were isolated from this patient (Wielders et al. 2001). The recent whole genome sequencing of both the *S. aureus* strains and the SCC*mec* elements has confirmed that the two *S. aureus* genome differed only by the presence of SCC*mec* element in MRSA and that the SCC*mec* of the MRSA and of the MR-S. *epidermidis* isolates differed only by a single nucleotide in the *ccrB* gene (Bloemendaal et al. 2010). However, SCC*mec* transfer could not be obtained *in vitro* by conjugation between these strains (Bloemendaal et al. 2010). The reason of this unsuccessfully *in vitro* transfer remains unclear.

The others clinical observations supported the hypothesis of transfer by showing high homology of the nucleotide sequence between SCC*mec* of MRSA and of MR-CoNS isolates in patients (Berglund and Soderquist 2008; Ibrahem et al. 2009; Jamaluddin et al. 2008; Wielders et al. 2001; Wisplinghoff et al. 2003; Zhang et al. 2009),

but they did not evidenced the transfer itself. However, we can speculate that the first step of the SCC*mec* transfer will consist in its excision from the donor strain driven by the *ccrAB* encoded recombinase, which is constitutive of all SCC*mec* (Katayama et al. 2000). The excision of SCC*mec* through CcrAB could then spontaneously occur, as show *in vitro* (Jansen et al. 2006; Noto and Archer 2006; Steidl et al. 2008). Antibiotics could play an important role in the transcription of *ccrA-ccrB* genes (Higgins et al. 2009; Steidl et al. 2008). Indeed, exposure to beta-lactam antibiotics has an upregulatory effect on the transcription of *ccrA* and vancomycin has an upregulatory effect on the transcription of *ccrA* of SCC*mec* IVa and an opposite effect on that of SCC*mec* II (Higgins et al. 2009). Thus, it has been hypothesized that vancomycin may have eased the spread of SCC*mec* IVa (Higgins et al. 2009). If the excision of SCC*mec* from the donor strain occurs, it remains to speculate how it is transferred to the recipient strain. It is known that *S. aureus* is not easily transformable. Also, transfer by conjugation has never been obtained (Lacey 1972). As for the transfer by transduction, it has been achieved successfully 30 years ago, but has never been reported since (Cohen and Sweeney 1973). If transduction was indeed the mechanisms of transfer, then SCC*mec* should be integrated in the chromosome. Recently, two new restrictions endonucleases, SauI type I restriction-modification system and one belonging to type III restriction-modification system have been shown to have an important role in the integration of foreign DNA in the chromosome of *S. aureus* (Corvaglia et al. 2010; Waldron and Lindsay 2006). It is interesting to note that their expression vary substantially between isolates which could explain why some STs could be more prone to accept SCC*mec* that others.

10.7 Conclusion

The first description of a MRSA strain has been reported in 1961, only 1 year after the use of the methicillin to treat infections caused by penicillin-resistant *S. aureus* strains. It has been necessary to wait more than 20 years to elucidate the mechanism which supports the methicillin resistance and 15 years more to discover that the *mecA* gene was on a mobile genetic element, SCC*mec*. During this time, MRSA has widely spread in most hospitals worldwide only and was mainly associated with healthcare. Recently, it has rapidly emerged in the community causing infections in individual without risk exposure to healthcare. Find what is the *mecA* and SCC*mec* reservoir for *S. aureus*, how SCC*mec* is acquired by *S. aureus*, and what are the factors which favour this transfer are important to develop strategies which reduces the *mecA* reservoir, and especially to control or to limit the transfer. Several lines of evidence support the hypothesis that MR-CoNS constitute a SCC*mec* reservoir for *S. aureus* such as the high prevalence of MR-CoNS nasal carriage in the community, just as in hospitalized patients, and the high diversity of SCC*mec* carried by MR-CoNS. The mechanism of transfer and the factors which cause the transfer still remained to be explored.

References

Abdel-Meguid SS, Grindley ND, Templeton NS et al (1984) Cleavage of the site-specific recombination protein gamma delta resolvase: the smaller of two fragments binds DNA specifically. Proc Natl Acad Sci USA 81:2001–2005

Anonymous (1999) Four pediatric deaths from community-acquired methicillin-resistant Staphylococcus aureus–Minnesota and North Dakota, 1997–1999. MMWR Morb Mortal Wkly Rep 48:707–710

Archer GL, Armstrong BC (1983) Alteration of staphylococcal flora in cardiac surgery patients receiving antibiotic prophylaxis. J Infect Dis 147:642–649

Armand-Lefevre L, Ruimy R, Andremont A (2005) Clonal comparison of Staphylococcus aureus isolates from healthy pig farmers, human controls, and pigs. Emerg Infect Dis 11:711–714

Barber M (1947) Staphylococcal infection due to penicillin-resistant strains. Br Med J 2:863–865

Barber M (1961) Methicillin-resistant staphylococci. J Clin Pathol 14:385–393

Barber M, Rozwadowska-Dowzenko M (1948) Infection by penicillin-resistant staphylococci. Lancet 2:641–644

Barbier F, Ruppe E, Hernandez D et al (2010) Methicillin-resistant coagulase-negative staphylococci in the community: high homology of SCCmec IVa between Staphylococcus epidermidis and major clones of methicillin-resistant Staphylococcus aureus. J Infect Dis 202:270–281

Bartels MD, Boye K, Rhod Larsen A et al (2007) Rapid increase of genetically diverse methicillin-resistant Staphylococcus aureus, Copenhagen, Denmark. Emerg Infect Dis 13:1533–1540

Becker K, Pagnier I, Schuhen B et al (2006) Does nasal cocolonization by methicillin-resistant coagulase-negative staphylococci and methicillin-susceptible Staphylococcus aureus strains occur frequently enough to represent a risk of false-positive methicillin-resistant S. aureus determinations by molecular methods? J Clin Microbiol 44:229–231

Berglund C, Soderquist B (2008) The origin of a methicillin-resistant Staphylococcus aureus isolate at a neonatal ward in Sweden-possible horizontal transfer of a staphylococcal cassette chromosome mec between methicillin-resistant Staphylococcus haemolyticus and Staphylococcus aureus. Clin Microbiol Infect 14:1048–1056

Bloemendaal AL, Brouwer EC, Fluit AC (2010) Methicillin resistance transfer from Staphylocccus epidermidis to methicillin-susceptible Staphylococcus aureus in a patient during antibiotic therapy. PLoS One 5:e11841

Bode LG, Kluytmans JA, Wertheim HF et al (2010) Preventing surgical-site infections in nasal carriers of Staphylococcus aureus. N Engl J Med 362:9–17

Chambers HF (1997) Methicillin resistance in staphylococci: molecular and biochemical basis and clinical implications. Clin Microbiol Rev 10:781–791

Chambers HF, Deleo FR (2009) Waves of resistance: Staphylococcus aureus in the antibiotic era. Nat Rev Microbiol 7:629–641

Cohen S, Sweeney HM (1973) Effect of the prophage and penicillinase plasmid of the recipient strain upon the transduction and the stability of methicillin resistance in Staphylococcus aureus. J Bacteriol 116:803–811

Coombs GW, Nimmo GR, Bell JM et al (2004) Genetic diversity among community methicillin-resistant Staphylococcus aureus strains causing outpatient infections in Australia. J Clin Microbiol 42:4735–4743

Corvaglia AR, Francois P, Hernandez D et al (2010) A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. Proc Natl Acad Sci USA 107:11954–11958

Costa SF, Miceli MH, Anaissie EJ (2004) Mucosa or skin as source of coagulase-negative staphylococcal bacteraemia? Lancet Infect Dis 4:278–286

Couto I, de Lencastre H, Severina E et al (1996) Ubiquitous presence of a mecA homologue in natural isolates of Staphylococcus sciuri. Microb Drug Resist 2:377–391

Couto I, Sanches IS, Sa-Leao R et al (2000) Molecular characterization of Staphylococcus sciuri strains isolated from humans. J Clin Microbiol 38:1136–1143

Couto I, Wu SW, Tomasz A et al (2003) Development of methicillin resistance in clinical isolates of Staphylococcus sciuri by transcriptional activation of the mecA homologue native to s. J Bacteriol 185:645–653

Cuny C, Friedrich A, Kozytska S et al (2010) Emergence of methicillin-resistant Staphylococcus aureus (MRSA) in different animal species. Int J Med Microbiol 300:109–117

Descloux S, Rossano A, Perreten V (2008) Characterization of new staphylococcal cassette chromosome mec (SCCmec) and topoisomerase genes in fluoroquinolone- and methicillin-resistant Staphylococcus pseudintermedius. J Clin Microbiol 46:1818–1823

Deurenberg RH, Stobberingh EE (2008) The evolution of Staphylococcus aureus. Infect Genet Evol 8:747–763

Enright MC, Day NP, Davies CE et al (2000) Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus aureus. J Clin Microbiol 38:1008–1015

Enright MC, Robinson DA, Randle G et al (2002) The evolutionary history of methicillin-resistant Staphylococcus aureus (MRSA). Proc Natl Acad Sci USA 99:7687–7692

Feil EJ, Li BC, Aanensen DM et al (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530

Gill SR, Fouts DE, Archer GL et al (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant Staphylococcus aureus strain and a biofilm-producing methicillin-resistant Staphylococcus epidermidis strain. J Bacteriol 187:2426–2438

Hanssen AM, Sollid JU (2007) Multiple staphylococcal cassette chromosomes and allelic variants of cassette chromosome recombinases in Staphylococcus aureus and coagulase-negative staphylococci from Norway. Antimicrob Agents Chemother 51:1671–1677

Hartman BJ, Tomasz A (1984) Low-affinity penicillin-binding protein associated with beta-lactam resistance in Staphylococcus aureus. J Bacteriol 158:513–516

Herold BC, Immergluck LC, Maranan MC et al (1998) Community-acquired methicillin-resistant Staphylococcus aureus in children with no identified predisposing risk. JAMA 279:593–598

Heusser R, Ender M, Berger-Bachi B et al (2007) Mosaic staphylococcal cassette chromosome mec containing two recombinase loci and a new mec complex, B2. Antimicrob Agents Chemother 51:390–393

Higashide M, Kuroda M, Omura CT et al (2008) Methicillin-resistant Staphylococcus saprophyticus isolates carrying staphylococcal cassette chromosome mec have emerged in urogenital tract infections. Antimicrob Agents Chemother 52:2061–2068

Higgins PG, Rosato AE, Seifert H et al (2009) Differential expression of ccrA in methicillin-resistant Staphylococcus aureus strains carrying staphylococcal cassette chromosome mec type II and IVa elements. Antimicrob Agents Chemother 53:4556–4558

Hiramatsu K, Cui L, Kuroda M et al (2001) The emergence and evolution of methicillin-resistant Staphylococcus aureus. Trends Microbiol 9:486–493

Huijsdens XW, van Dijke BJ, Spalburg E et al (2006) Community-acquired MRSA and pig-farming. Ann Clin Microbiol Antimicrob 5:26

Ibrahem S, Salmenlinna S, Virolainen A et al (2009) Carriage of methicillin-resistant Staphylococci and their SCCmec types in a long-term-care facility. J Clin Microbiol 47:32–37

International Working Group on the Classification of Staphylococcal Cassette Chromosome Elements (2009) Classification of staphylococcal cassette chromosome mec (SCCmec): guidelines for reporting novel SCCmec elements. Antimicrob Agents Chemother 53:4961–4967

Ito T, Katayama Y, Hiramatsu K (1999) Cloning and nucleotide sequence determination of the entire mec DNA of pre-methicillin-resistant Staphylococcus aureus N315. Antimicrob Agents Chemother 43:1449–1458

Ito T, Katayama Y, Asada K et al (2001) Structural comparison of three types of staphylococcal cassette chromosome mec integrated in the chromosome in methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother 45:1323–1336

Ito T, Ma XX, Takeuchi F et al (2004) Novel type V staphylococcal cassette chromosome mec driven by a novel cassette chromosome recombinase, ccrC. Antimicrob Agents Chemother 48:2637–2651

Jamaluddin TZ, Kuwahara-Arai K, Hisata K et al (2008) Extreme genetic diversity of methicillin-resistant Staphylococcus epidermidis strains disseminated among healthy Japanese children. J Clin Microbiol 46:3778–3783

Jansen WT, Beitsma MM, Koeman CJ et al (2006) Novel mobile variants of staphylococcal cassette chromosome mec in Staphylococcus aureus. Antimicrob Agents Chemother 50:2072–2078

Jarlier V, Trystram D, Brun-Buisson C et al (2010) Curbing methicillin-resistant Staphylococcus aureus in 38 French hospitals through a 15-year institutional control program. Arch Intern Med 170:552–559

Jevons MP (1961) "Celbenin" resistant staphylococci. Br Med J 1:124–125

Katayama Y, Ito T, Hiramatsu K (2000) A new class of genetic element, staphylococcus cassette chromosome mec, encodes methicillin resistance in Staphylococcus aureus. Antimicrob Agents Chemother 44:1549–1555

Katayama Y, Ito T, Hiramatsu K (2001) Genetic organization of the chromosome region surrounding mecA in clinical staphylococcal strains: role of IS431-mediated mecI deletion in expression of resistance in mecA-carrying, low-level methicillin-resistant Staphylococcus haemolyticus. Antimicrob Agents Chemother 45:1955–1963

Kirby WM (1944) Extraction of a highly potent penicillin inactivator from penicillin resistant Staphylococci. Science 99:452–453

Kloos W (1997) Taxonomy and systematics of staphylococci indigenous to humans. In: Crossley KB, Archer G (eds) The Staphylococci in human Disease. Churchill Livingstone, New York, pp 113–137

Kloos WE, Musselwhite MS (1975) Distribution and persistence of Staphylococcus and Micrococcus species and other aerobic bacteria on human skin. Appl Microbiol 30:381–385

Kloos WE, Ballard DN, Webster JA et al (1997) Ribotype delineation and description of Staphylococcus sciuri subspecies and their potential as reservoirs of methicillin resistance and staphylolytic enzyme genes. Int J Syst Bacteriol 47:313–323

Kobayashi N, Alam MM, Urasawa S (2001) Genomic rearrangement of the mec regulator region mediated by insertion of IS431 in methicillin-resistant staphylococci. Antimicrob Agents Chemother 45:335–338

Kondo Y, Ito T, Ma XX et al (2007) Combination of multiplex PCRs for staphylococcal cassette chromosome mec type assignment: rapid identification system for mec, ccr, and major differences in junkyard regions. Antimicrob Agents Chemother 51:264–274

Krediet TG, Jones ME, Janssen K et al (2001) Prevalence of molecular types and mecA gene carriage of coagulase-negative Staphylococci in a neonatal intensive care unit: relation to nosocomial septicemia. J Clin Microbiol 39:3376–3378

Lacey R (1972) Transfer of chromosomal genes between staphylococci in mixed culture. J Gen Microbiol 71:399–401

Lee YL, Cesario T, Tran C et al (2000) Nasal colonization by methicillin-resistant coagulase-negative staphylococcus in community skilled nursing facility patients. Am J Infect Control 28:269–272

Lowy FD (1998) Staphylococcus aureus infections. N Engl J Med 339:520–532

Lucet JC, Chevret S, Durand-Zaleski I et al (2003) Prevalence and risk factors for carriage of methicillin-resistant Staphylococcus aureus at admission to the intensive care unit: results of a multicenter study. Arch Intern Med 163:181–188

Ma XX, Ito T, Tiensasitorn C et al (2002) Novel type of staphylococcal cassette chromosome mec identified in community-acquired methicillin-resistant Staphylococcus aureus strains. Antimicrob Agents Chemother 46:1147–1152

Matsuhashi M, Song MD, Ishino F et al (1986) Molecular cloning of the gene of a penicillin-binding protein supposed to cause high resistance to beta-lactam antibiotics in Staphylococcus aureus. J Bacteriol 167:975–980

Miragaia M, Couto I, de Lencastre H (2005) Genetic diversity among methicillin-resistant Staphylococcus epidermidis (MRSE). Microb Drug Resist 11:83–93
Miragaia M, Carrico JA, Thomas JC et al (2008) Comparison of molecular typing methods for characterization of Staphylococcus epidermidis: proposal for clone definition. J Clin Microbiol 46:118–129
Moran GJ, Krishnadasan A, Gorwitz RJ et al (2006) Methicillin-resistant S. aureus infections among patients in the emergency department. N Engl J Med 355:666–674
Naimi TS, LeDell KH, Boxrud DJ et al (2001) Epidemiology and clonality of community-acquired methicillin-resistant Staphylococcus aureus in Minnesota, 1996–1998. Clin Infect Dis 33:990–996
Noto MJ, Archer GL (2006) A subset of Staphylococcus aureus strains harboring staphylococcal cassette chromosome mec (SCCmec) type IV is deficient in CcrAB-mediated SCCmec excision. Antimicrob Agents Chemother 50:2782–2788
Nubel U, Roumagnac P, Feldkamp M et al (2008) Frequent emergence and limited geographic dispersal of methicillin-resistant Staphylococcus aureus. Proc Natl Acad Sci USA 105:14130–14135
O'Brien FG, Lim TT, Chong FN et al (2004) Diversity among community isolates of methicillin-resistant Staphylococcus aureus in Australia. J Clin Microbiol 42:3185–3190
Okuma K, Iwakawa K, Turnidge JD et al (2002) Dissemination of new methicillin-resistant Staphylococcus aureus clones in the community. J Clin Microbiol 40:4289–4294
Oliveira DC, Tomasz A, de Lencastre H (2002) Secrets of success of a human pathogen: molecular evolution of pandemic clones of meticillin-resistant Staphylococcus aureus. Lancet Infect Dis 2:180–189
Oliveira DC, Milheirico C, de Lencastre H (2006) Redefining a structural variant of staphylococcal cassette chromosome mec, SCCmec type VI. Antimicrob Agents Chemother 50:3457–3459
Perl TM, Cullen JJ, Wenzel RP et al (2002) Intranasal mupirocin to prevent postoperative Staphylococcus aureus infections. N Engl J Med 346:1871–1877
Robinson DA, Enright MC (2003) Evolutionary models of the emergence of methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother 47:3926–3934
Ruppe E, Barbier F, Mesli Y et al (2009) Diversity of staphylococcal cassette chromosome mec structures in methicillin-resistant Staphylococcus epidermidis and Staphylococcus haemolyticus strains among outpatients from four countries. Antimicrob Agents Chemother 53:442–449
Schnellmann C, Gerber V, Rossano A et al (2006) Presence of new mecA and mph(C) variants conferring antibiotic resistance in Staphylococcus spp. isolated from the skin of horses before and after clinic admission. J Clin Microbiol 44:4444–4454
Silva FR, Mattos EM, Coimbra MV et al (2001) Isolation and molecular characterization of methicillin-resistant coagulase-negative staphylococci from nasal flora of healthy humans at three community institutions in Rio de Janeiro City. Epidemiol Infect 127:57–62
Soderquist B, Berglund C (2009) Methicillin-resistant Staphylococcus saprophyticus in Sweden carries various types of staphylococcal cassette chromosome mec (SCCmec). Clin Microbiol Infect 15:1176–1178
Steidl R, Pearson S, Stephenson RE et al (2008) Staphylococcus aureus cell wall stress stimulon gene-lacZ fusion strains: potential for use in screening for cell wall-active antimicrobials. Antimicrob Agents Chemother 52:2923–2925
Stephens AJ, Huygens F, Giffard PM (2007) Systematic derivation of marker sets for staphylococcal cassette chromosome mec typing. Antimicrob Agents Chemother 51:2954–2964
Suzuki E, Kuwahara-Arai K, Richardson JF et al (1993) Distribution of mec regulator genes in methicillin-resistant Staphylococcus clinical strains. Antimicrob Agents Chemother 37:1219–1226
Terpstra S, Noordhoek GT, Voesten HG et al (1999) Rapid emergence of resistant coagulase-negative staphylococci on the skin after antibiotic prophylaxis. J Hosp Infect 43:195–202
Tristan A, Bes M, Meugnier H et al (2007) Global distribution of Panton-Valentine leukocidin–positive methicillin-resistant Staphylococcus aureus, 2006. Emerg Infect Dis 13:594–600
Tsubakishita S, Kuwahara-Arai K, Sasaki T et al (2010) Origin and molecular evolution of the determinant of methicillin resistance in staphylococci. Antimicrob Agents Chemother 54:4352–4359

Ubukata K, Yamashita N, Konno M (1985) Occurrence of a beta-lactam-inducible penicillin-binding protein in methicillin-resistant staphylococci. Antimicrob Agents Chemother 27:851–857

Udo EE, Pearman JW, Grubb WB (1993) Genetic analysis of community isolates of methicillin-resistant Staphylococcus aureus in Western Australia. J Hosp Infect 25:97–108

Vandenesch F, Naimi T, Enright MC et al (2003) Community-acquired methicillin-resistant Staphylococcus aureus carrying Panton-Valentine leukocidin genes: worldwide emergence. Emerg Infect Dis 9:978–984

von Eiff C, Becker K, Machka K et al (2001) Nasal carriage as a source of Staphylococcus aureus bacteremia. Study Group. N Engl J Med 344:11–16

Waldron DE, Lindsay JA (2006) Sau1: a novel lineage-specific type I restriction-modification system that blocks horizontal gene transfer into Staphylococcus aureus and between S. aureus isolates of different lineages. J Bacteriol 188:5578–5585

Wertheim HF, Vos MC, Ott A et al (2004) Risk and outcome of nosocomial Staphylococcus aureus bacteraemia in nasal carriers versus non-carriers. Lancet 364:703–705

Wertheim HF, Melles DC, Vos MC et al (2005) The role of nasal carriage in Staphylococcus aureus infections. Lancet Infect Dis 5:751–762

Wielders CL, Vriens MR, Brisse S et al (2001) In-vivo transfer of mecA DNA to Staphylococcus aureus [corrected]. Lancet 357:1674–1675

Wisplinghoff H, Rosato AE, Enright MC et al (2003) Related clones containing SCCmec type IV predominate among clinically significant Staphylococcus epidermidis isolates. Antimicrob Agents Chemother 47:3574–3579

Wu S, Piscitelli C, de Lencastre H et al (1996) Tracking the evolutionary origin of the methicillin resistance gene: cloning and sequencing of a homologue of mecA from a methicillin susceptible strain of Staphylococcus sciuri. Microb Drug Resist 2:435–441

Zhang YQ, Ren SX, Li HL et al (2003) Genome-based analysis of virulence genes in a non-biofilm-forming Staphylococcus epidermidis strain (ATCC 12228). Mol Microbiol 49:1577–1593

Zhang K, McClure JA, Elsayed S et al (2009) Novel staphylococcal cassette chromosome mec type, tentatively designated type VIII, harboring class A mec and type 4 ccr gene complexes in a Canadian epidemic strain of methicillin-resistant Staphylococcus aureus. Antimicrob Agents Chemother 53:531–540

Chapter 11
Molecular Knowledge of Mechanisms of Helminth Resistance: Importance for Diagnostic and Epidemiology

Anne Silvestre and Jacques Cabaret

Abstract Helminths infestations in animals and humans are mostly controlled by anthelmintics. Large-scale treatment has led to resistance to many drugs, namely benzimidazoles, avermectins and levamisole. Whereas avermectins and levamisole resistance appear to be multigenic, benzimidazole resistance is largely monogenic, based on polymorphism in β-tubulin gene. Spatial distribution of β-tubulin alleles in closed helminths populations supports a common origin of alleles on a specific farm. The main forces responsible for anthelmintic resistance development in field populations are the introduction of ancestral alleles (i.e. pre existing polymorphism in helminths before herds' constitution) and the selection of alleles appeared after herds' constitution. The acquisition of resistance may have a cost that counterbalance the advantage of being resistant. For benzimidazole resistance, the interaction between life-traits combining advantages and disadvantages for resistant genotypes, compared to susceptible ones, may explain the stability of acquired resistance along years. Much remains to be done for avermectins and levamisole resistance, for which molecular mechanisms remain to be identified.

11.1 Introduction

Helminths are widespread in animals and humans. Control is mostly based on the use of anthelmintics at a very large scale in veterinary and human medicines. The large use of drugs and appearance of resistance to them has been recorded in many other pathogens than helminths (bacteria, fungi and protozoa) and seems to be a general rule. The large-scale treatment or prophylaxis has led to resistance to many of the

A. Silvestre (✉) • J. Cabaret
INRA, UR1282 Infectiologie Animale et Santé Publique, F-37380 Nouzilly, France
e-mail: anne.silvestre@tours.inra.fr; jacques.cabaret@tours.inra.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_11,

front-line drugs against malaria in man (Enyati and Hemingway 2010) and against helminths (mostly nematodes) in animals (Kaplan 2004; Silvestre et al. 2002). Helminth parasite control in human (Pion et al. 2009) and animal health is threatened by the increasing development of anthelmintic resistance against the main drug families, namely benzimidazoles (BZ), avermectins and levamisole (Kaplan 2004; Pomroy 2006). Cases of multi-resistance (i.e. resistance against BZ and ivermectin or levamisole) are appearing in many areas of the world (Bartley et al. 2004; Besier et al. 2003; Sargison et al. 2007; Vermunt et al. 1995). Anthelmintic resistance is a major topic of helminth epidemiology research as it is for antimicrobial or virus drug resistant pathogens (Riley 2004). The populations of helminths are then distributed into distinct groups of different resistance status and hence will have different epidemiological traits.

Genetic mechanisms conferring resistance are many and include detoxication (MDR Multidrug resistance, (Kerboeuf et al. 2003)), loss of drug binding, modification of response after the drug bound its target and loss of the drug target. Whereas avermectins and levamisole resistance appear to be multigenic, BZ resistance which is the best characterized mechanism is largely monogenic: it is conferred by amino acid substitutions in β-tubulin gene that reduce affinity binding of BZ molecules for β-tubulin (Kwa et al. 1994; Lacey and Prichard 1986; Silvestre and Cabaret 2002) (Table 11.1). Thus the fragmentation of populations into groups of different level of resistance will be restricted in BZ-resistance but probably fragmentation for other drugs may be more important (into highly, medium and low resistant for example).

11.2 Benzimidazole Resistance: From Well Known Genetic Mechanism to Epidemiology of Resistance

11.2.1 From Genetic Mechanisms of Resistance to Diagnostic

Several mutations in β-tubulin gene may confer BZ resistance, among which F200Y plays a major role (see Box 11.1). Based on the identification of those molecular mechanisms, several diagnostic tools were developed: on individual larvae (Silvestre and Humbert 2000) and on individual adult worms, with PCR-RFLP (Shayan et al. 2007; Tiwari et al. 2006). More recently, real time PCR offered a rapid method for detection of BZ resistance allele on multiple worms, with a reliable detection threshold of 5% (von Samson-Himmelstjerna et al. 2009; Walsh et al. 2007). Those molecular tools may not be used for routine molecular resistance testing on farms for small ruminants (BZ resistance frequency is too high), but they have proved very useful to evaluate the sustainability of management strategies (Leignel and Cabaret 2001; Leignel et al. 2010). They also may help to monitor the early drug resistance development in cattle farms (Coles 2002). Identification of drug target and development of reliable molecular diagnoses are prerequisites for the study of mechanisms of resistance gene diffusion.

Table 11.1 Target sites and molecular mechanism of resistance for commonly used anthelmintics

Anthelmintic	Benzimidazoles	Macrocyclic lactones	Levamisole	Monepantel	Praziquantel
Target site	Beta-tubulin	Glutamate-gated chloride-channel	Nicotinic acetylcholine receptor	Nicotinic acetylcholine receptor	Voltage-gated Ca^{2+} channels
SNP associated to resistance	F167Y (*isotype 1*) E198A (*isotype 1*) F200Y (*isotype 1*)	A169V (*Hco-glc-5*) K169R (*Hco-lgc-37*) F256L (*avr-14*)	No SNP known	Mis-splicing (*Hco-mptl-1, Hco-des-2*)	S250C (*β-subunit*) S260A (*β-subunit*)
Efflux	–	P-glycoprotein over-expression	–	–	–

Box 11.1 Genetic Mechanisms of Benzimidazole Resistance

A large number of laboratory strains as well as field isolates were studied for BZ resistance. The most frequently detected change associated with BZ resistance is a point mutation in codon 200 of the isotype 1 β-tubulin gene. This mutation leads to the replacement of phenylalanine by tyrosine (F200Y) (Beech et al. 1994; Elard et al. 1996; Elard and Humbert 1999; Kwa et al. 1994; Palcy et al. 2010; Silvestre and Cabaret 2002). The functional role of this F200Y substitution was demonstrated by directed mutagenesis of β -tubulin in the free living nematode *Caenorhabditis elegans* (Kwa et al. 1995): F200Y substitution causes a loss of binding of BZ and allows tubulin to polymerize in the presence of BZ. Increasing number of studies is conducted on field isolates, allowing the identification of rare and new polymorphisms conferring BZ resistance. The substitution of phenylalanine for tyrosine at position 167 (F167Y) confers BZ resistance to *Teladorsagia circumcincta* and *H. contortus* (Prichard et al. 2000; Silvestre and Cabaret 2002). The replacement of glutamate by alanine at position 198 (E198A) was found in BZ resistant isolates of *H. contortus* (Rufener et al. 2009). A field isolate of *T. circumcincta*, which segregates for all three β -tubulin resistance mutations, allowed the study of possible interactions between them. Each mutation is recessive and only homozygous mutants survived BZ treatments (Elard and Humbert 1999; Silvestre and Cabaret 2002). Complementation occurs between F200Y and F167Y in that double heterozygotes survive BZ treatment just as well as individuals homozygous for F200Y alone. This is also true for complementation between F200Y and E198A resistance mutations (Silvestre, unpub results).

11.2.2 Diffusion of Resistance Alleles

Two isotype 1 β-tubulin types (Type I and Type II) were found (Leignel et al. 2002; Silvestre and Humbert 2002) after sequencing of a central part of β-tubulin gene from *Teladorsagia circumcincta* field isolates from central and south-western France. Each isotype was characterised by synonymous SNP (single nucleotide polymorphism) in coding regions, three indels and many SNP in introns. Both types comprised BZ–resistant sequences with a F200Y SNP (Elard 1998) indicating that BZ-resistant sequences were not monophyletic; in other words, SNP F200Y appeared independently in several populations. No data is available for the other main species, *Haemonchus contortus* and *Trichostrongylus colubriformis*. These multiple independent origins of BZ-resistant alleles are also supported by the diversity of SNP conferring BZ-resistance (namely, F200Y, E198A and F167Y).

Table 11.2 Distribution of β-tubulin alleles in the *Teladorsagia circumcincta* (TCR) and *Haemonchus contortus* (HCR) populations. Alleles were sequenced from 5 to 15 individual adult male nematodes (Data from Silvestre and Humbert 2002)

Farm	*T. circumcincta*	*H. contortus*
BER	–	HCR5
BON	–	HCR3
DEL	TCR5	–
FRA	TCR3 TCR5	–
GAR	TCR8	–
GRO	–	HCR4 HCR6
HUM	–	HCR3
PIC	TCR3	HCR1
SER	TCR7	HCR1
SOR	TCR3 TCR4	HCR1 HCR2

Based on isotype 1 β-tubulin gene sequencing, the spatial distribution of BZ-resistant alleles between field isolates was investigated (Silvestre and Humbert 2002) (Table 11.2). It was found that *T. circumcincta* and *H. contortus* had the same allelic diversity (eight alleles in ten populations and six alleles in seven populations, respectively). In *T. circumcincta*, some private alleles (TCR4, TCR5 and TCR8) were found (present in only one population) and some alleles were shared by 2–3 populations (TCR3 and TCR7). Similar findings were obtained for *H. contortus* (HCR1 as shared allele by several populations, and HCR2, HCR3, HCR4, HCR5 and HCR6 as private alleles).

To determine whether if shared alleles were identical by descent (due to migration or to introduction of ancestral allele) or by homoplasy (independent appearance of the same alleles in different populations), the genetic relatedness of BZ-resistant populations was studied. Direct study of dispersal and population structure of helminths parasites are impeded by their parasitic development inside hosts. Consequently, the use of polymorphic molecular markers and their variation within and between predefined populations allows genetic epidemiology of helminths parasites. Historical gene flows between populations are rarely known, and inferences can be made from neutral genetic markers (i.e. not submitted to selective forces). Among genetic markers, mitochondrial DNA (mtDNA) evolves rapidly and lacks recombination. Based on ND4 gene polymorphism from mtDNA, it was demonstrated that population structure of helminths parasites is highly dependant on population structure of their hosts (Blouin et al. 1995). *T. circumcincta* field isolates were characterised with mtDNA: no genetic subdivision was observed on a region scale (200 Km) but significant genetic subdivision was observed on larger geographical scale, between regions (Leignel and Humbert 2001).

Mitochondrial DNA is uniparentally inherited and coding sequences may not be neutral (i.e. not only affected by demographic and geographical processes). Other genetic markers were developed: microsatellites (MS) are short tandemly repeated sequences of DNA of 2–3 base pairs. Polymorphism of MS markers consists of variation in the number of repeats of the sequence. MS are highly polymorphic,

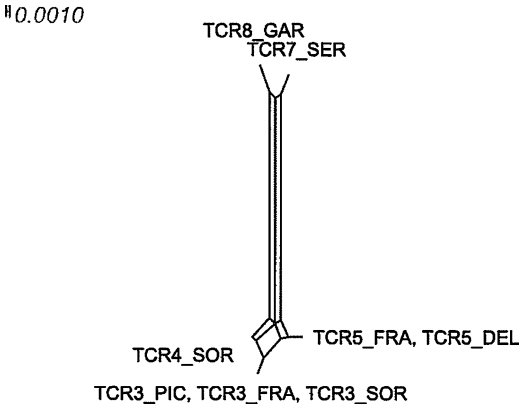

Fig. 11.1 Network built with isotype 1 β-tubulin sequences from the nematode *Teladorsagia circumcincta*. Only one sequence is shown for each allele and each population (i.e. each farm). Alleles were sequenced from 5 to 15 individual adult male nematodes, data from (Silvestre and Humbert 2002). SplitsTree v4.8 software (Huson 1998) was used to build the network with the Neighbour-net method, a distance-based method, with the idea of minimizing the distances (or number of mutations) within populations (or haplotypes). Each sequence of β-tubulin is coded as follows: 'TC' refers to *T. circumcincta*, "R" to benzimidazole resistant sequence (Phe200Tyr polymorphism), the figure identifies the allele and the three following letters refer to the farm name

co-dominant (allowing all homozygous and heterozygous genotypes distinction) and abundant throughout the genome and were developed for *T. circumcincta* and for *H. contortus* (Grillo et al. 2006; Hoekstra et al. 1997; Otsen et al. 2000). Based on MS polymorphism, genetic relatedness between populations was characterised. No genetic subdivision was observed between *T. circumcincta* populations from a given region, but the genetic subdivision was observed between populations from distinct regions, confirming mtDNA data (Leignel and Humbert 2001). Conversely, *H. contortus* populations from a given region displayed marked genetic differentiation, suggesting that severe bottlenecks occurred in this species, after farms had been established (Silvestre et al. 2009). Life trait history of both species may explain these discrepancies: *H. contortus* populations are strongly reduced during unfavourable periods and their high reproductive rate allows the species to persist even after severe population reduction. Conversely, *T. circumcincta* is well adapted to temperate climate, and is very abundant throughout the year. Absence of genetic subdivision may be due to consistently large effective population (as populations were isolated, large gene flow between population is irrelevant).

In this study, the sampled herds are constituted from animals from other local farms that have ceased operating, no infected host was introduced since the farms had been established up to 20 years previously (Silvestre et al. 2009). As a consequence of this local recruitment of the herds, each population may be genetically related to several other ones in the same region. Phylogeny based on network is more relevant than phylogeny based on bifurcating trees to illustrate the multiple connections between sampled populations. Network built with β-tubulin sequences from *T. circumcincta* (Fig. 11.1) indicates that TCR3 is the most abundant allele,

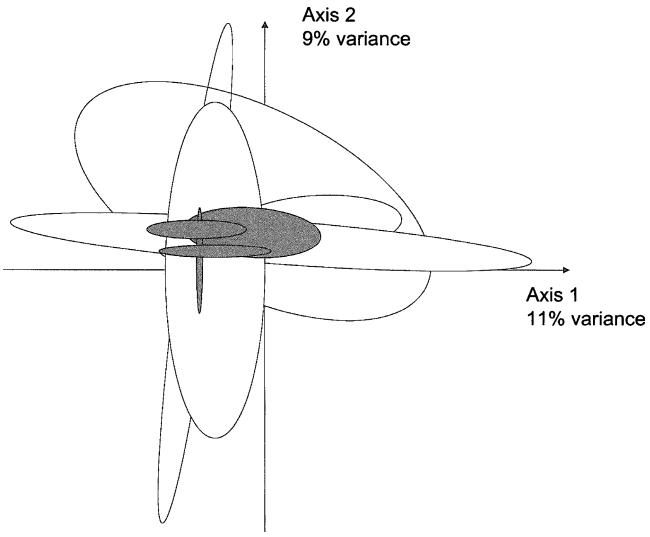

Fig. 11.2 Restriction fragment length polymorphism analysis with five enzymes, of the isotype 1 β-tubulin gene from five benzimidazole-susceptible (*unshaded ellipses*) and four–resistant natural *Teladorsagia circumcincta* isolates (*shaded ellipses*). Correspondence analysis (Adapted from Elard and Humbert (1999))

present in three oldest populations (constituted 26–30 years ago), supporting its ancestral origin. Alleles TCR4 and TCR5 recently diverged from this ancestral allele. The presence of TCR3 and TCR4 in SOR farm and the presence of TCR3 and TCR5 in FRA farm support this interpretation. Alleles TCR7 and TCR8 are poorly related to allele TCR3, supporting a different origin of these sequences. SER farm was established earlier than GAR farm, supporting that TCR7 may be an ancestral allele, and that TCR8 has recently diverged from TCR7 allele. TCR3 and TCR7 are thought to be "ancestral" alleles because at the time when the farms originated (the late 1970s, early 1980s), BZ resistance was infrequent in field populations, and BZ-resistant alleles may be extremely rare. They can be called "pre existing alleles" as (Elard and Humbert 1999) showed that BZ treatment reduced genetic variability of isotype 1 β-tubulin gene (estimated by restriction polymorphism) in *T. circumcincta* populations (Fig. 11.2). TCR3 and TCR7 should be ancestral alleles that were introduced in farms when herds were constituted, because populations have been isolated since then. But, the lack of genetic subdivision between *T. circumcincta* populations indicates that another hypothesis is possible: in different populations that share the same haplotypes, the F200Y SNP may arise several times, either soon after isolation or later on (Gilleard and Beech 2007). To test these hypotheses, the study of other species than *T. circumcincta* may be useful. If ancestral alleles from *T. circumcincta* were introduced when herds were constituted, we should expect that other species may also have been introduced. Consequently, other species should have shared alleles. This is the case for SOR and PIC farms: they share

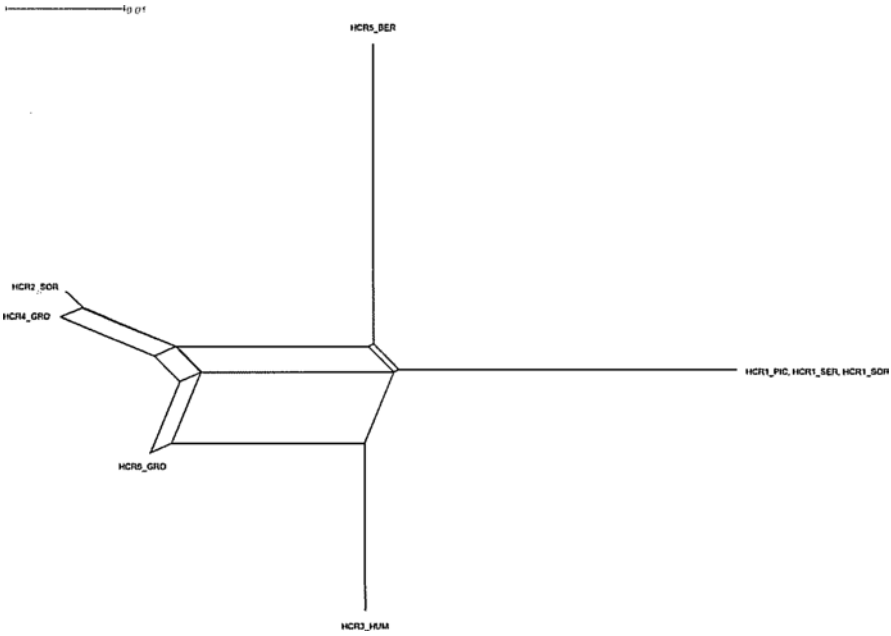

Fig. 11.3 Network built with isotype 1 β-tubulin sequences from the nematode *Haemonchus contortus*. Only one sequence is shown for each allele and each population (i.e. each farm). Alleles were sequenced from 5 to 15 individual adult male nematodes, data from (Silvestre and Humbert 2002). SplitsTree v4.8 software (Huson 1998) was used to build the network with the Neighbour-net method, a distance-based method, with the idea of minimizing the distances (or number of mutations) within populations (or haplotypes). Each sequence of β-tubulin is coded as follows: 'HC' refers to *H. contortus*, "R" to benzimidazole resistant sequence (Phe200Tyr polymorphism), the figure identifies the allele and the three following letters refer to the farm name

TCR3 for *T. circumcincta*, HCR1 for *H. contortus* (Fig. 11.3) and also the same BZ-allele for *Trichostrongylus colubriformis* (the third main species found) (Silvestre and Humbert 2002) (Table 11.2).

Network built with β-tubulin sequences from *H. contortus* (Fig. 11.3) indicates that HCR1, the most abundant allele, is present in three oldest populations (constituted 26–30 years ago), supporting its ancestral origin. The high frequency of HCR1 among individuals and among populations supports its ancestral status. Proximity of β-tubulin allele sequences of HCR2, HCR3, HCR4, HCR5 and HCR6 indicates that they recently diverged from the ancestral allele HCR1. The presence of HCR1 and HCR2 in SOR farm supports this interpretation. Private alleles, HCR2, HCR3, HCR4, HCR5 and HCR6, support the hypothesis that new alleles were generated since the herds have long been isolated.

As a conclusion, the introduction of ancestral alleles and the selection of alleles appeared after herds' constitution were responsible for the existence of several BZ-resistant alleles in *T. circumcincta* and *H. contortus* populations. The spatial

distribution of private alleles clearly matches the geographical distribution of the farms, indicating a common origin of alleles on a specific farm. The sampling design is crucial in molecular epidemiology studies. In the present studies (Leignel and Humbert 2001; Silvestre et al. 2009), field isolates were sampled from closed farms to be sure that migration was negligible, and that only introduction of alleles at the constitution of farms may account for gene flow between populations. BZ-resistance is an extremely useful model to our knowledge of drug resistance acquisition and diffusion. Thanks to the development of neutral genetic markers, with the identification of β-tubulin as BZ target, the processes involved in epidemiology and evolutionary genetics of anthelmintic resistance are better known. Much remains to be done for levamisole and avermectins resistance, for which molecular resistance mechanisms remain to be identified.

11.2.3 Does Fitness Modulate the Spread of BZ Genetic Resistance in Nematodes?

The acquisition of resistance may have a cost that counterbalance the advantage of being resistant. The resistance of the trematode *Schistosoma mansoni* to praziquantel is indicative of several outcomes in relation to resistance, some resistant isolates presenting reversion to susceptibility and other not (William et al. 2001). Comparative fitness studies of resistant versus susceptible worms are very few, outcome of resistance selection may be varied, and thus conclusions on the role of fitness are yet limited. In short term experimental conditions, no differences was evidenced in fitness of *T. circumcincta*, considering survival and egg production (Elard et al. 1998). No reversion was observed for the same species in untreated sheep during two grazing seasons (Leignel et al. 2010), the proportion of resistant worms remaining similar (circa 25%). However, susceptible worms submitted to increasing selective pressure by anthelmintics increased in size (by 6–10%); it was hypothesized that this was partly under the control of sheep, as treated lambs may mount and maintain a better response when infected (premunition) (Leignel and Cabaret 2001). The size of nematodes is an operational indicator of fecundity and it could be an advantage for resistant worms but since no reversion or increase of resistance was observed during the two grazing seasons, we have to hypothesize that other life-traits may be altered negatively. Several life-traits were studied in *Haemonchus contortus* (Melo 2005) and resistant worms compared with susceptible ones had higher establishment rate, higher egg excretion, these eggs developed better at lower temperature into infective larvae, and induced less anaemia in lambs, but the adult stage survival seemed shorter. In Brazilian conditions of Ceara State, we may suggest that resistant worms may present advantage during the rainy season (due to higher development into infective stage at lower temperature) which may be counterbalanced by their shorter life-span. The interactions between life-traits combining advantages and disadvantages for resistant worms compared susceptible ones, may explain the relative stability of acquired resistance along the years. Nevertheless

fitness of worms is only one component acting on maintenance or increase of the proportion of resistant worms. Host management plays also an important role in BZ resistance (Silvestre et al. 2002) or macrocyclic lactones (Lawrence et al. 2006), particularly the repeated use of the same drug (acting on selection), the presence of untreated refugia (acting on the maintenance of susceptibility alleles) and the importing of resistant parasites with purchased hosts.

11.3 Towards Epidemiology of Levamisole Resistance in Nematodes

Levamisole (LEV) is a cholinergic agonist widely used to control parasitic nematodes of humans and animals. It induces hyper contraction of body wall muscle, leading to spastic paralysis (Robertson and Martin 1993), and subsequent expulsion of worms by means of gastrointestinal transit. However, LEV efficacy has been compromised by the emergence of resistant parasites, specifically in gastrointestinal nematodes of small ruminants (Kaplan 2004). LEV resistance is thought to be associated with changes in binding characteristics or in the number of L-AChR channels present (Sangster et al. 1998). The pharmacological target of LEV and other cholinergic agonists such as pyrantel and tribendimidine (Kaminsky et al. 2008) are acetyl-choline receptors (AChRs) that mediate fast synaptic transmission at the neuro–muscular junction of nematodes. These receptors can be classified as responding to LEV, nicotine or bephenium and each has distinct electrophysiological properties (Hu et al. 2009; Martin and Robertson 2007). Molecular characterization of the levamisole AChR (L-AChR) was first achieved in the free-living nematode *Caenorhabditis elegans* (Boulin et al. 2008). Forward genetic screens have been extremely productive for identifying genes that encode L-AChR subunits and other factors required for their assembly (Jones et al. 2005). Five AChR subunit genes were identified in LEV resistant *C. elegans*: unc-38, unc-63, lev-8, unc-29 and lev-1 (Fleming et al. 1997; Towers et al. 2005). Two AChR subunits encoded by the orthologous of unc-38 and unc-29 genes are necessary in *Ascaris suum* (swine roundworm) to constitute a receptor sensitive to LEV when co-expressed in *Xenopus* oocytes (Williamson et al. 2009). These findings highlight the diversity of nematode L-AChR. The resistance to LEV of *Oesophogostomum dentatum* (pig roundworm) has been characterized as the loss of the specific LEV receptor (Robertson et al. 1999). The resistance to pyrantel (another cholinergic agonist) in *Ancylostoma caninum* (dog hookworm) is associated with reduced expression of genes orthologous to unc-38, unc-63 and unc-29 (Kopp et al. 2009). Complete coding sequences of the L-AChR subunits in two LEV-resistant and three susceptible isolates of *H. contortus, T. circumcincta* and *T. colubriformis* were essentially unchanged, but abbreviated transcripts of the unc-63 subunit were specifically expressed in resistant isolates of all three species (Neveu et al. 2007, 2010). Thus genetic of resistance to levamisole remains not fully understood and molecular epidemiology of resistance is still to be developed.

11.4 Resistance to Macrocyclic Lactones: Deciphering the Mechanisms

Macrocyclic lactones (ML, ivermectin and moxidectin for instance) bind to glutamate-gated chloride-channel irreversibly and cause a flaccid paralysis of helminths (Cully and Paress 1991; Cully et al. 1994; Dent et al. 2000; Geary et al. 1993); due to gastrointestinal motility the worms are expulsed from hosts. Three genes avr-14, avr-15 and glc-1 encode for subunit of this glutamate-gated chloride-channel (Dent et al. 2000). In the gastrointestinal nematode *Haemonchus contortus*, some polymorphisms in chloride-channel subunits were found in ML-resistant isolates. The mutation Ala169Val in *Hco-glc-5* and Lys169Arg in Hco-glc-37 observed in ML-resistant laboratory isolates may alter the normal function of the channel, modifying the transduction of the structural changes induced by ligand binding with the chloride-channel (Taly et al. 2009). Feeding behaviour and movement of adult *H. contortus* with or without this polymorphism give more evidence to support a functional link between polymorphism in amino acid 169 and ML-resistance. Frequencies of these alleles were significantly higher in ML-resistant laboratory isolates in comparison with ML-susceptible ones. However, these mutations have not been observed in field isolates (Beech and Silvestre 2010). In *H. contortus* and another gastrointestinal nematode species, *Cooperia oncophora*, a Phe256Leu polymorphism was observed in avr-14 gene and decreases sensitivity to ivermectin (McCavera et al. 2009; Njue and Prichard 2004). Here again, the mutation Phe256Leu may affect the transduction of structural changes induced by ligand binding with the chloride-channel, but its frequency in only moderately increased in ML-resistant isolates.

In ML-resistance, active drug efflux by multi-drug resistance (MDR) mechanisms has been observed (Bourguinat et al. 2008): in an ivermectin-selected *H. contortus* strain, PgpA (P-glycoprotein A) was over-expressed in comparison with the parental unselected strain (Xu et al. 1998). Furthermore, the use of the MDR-reversing agents (i.e. verapamil and CL347099) increased the efficacy of ivermectin and moxidectin against resistant *H. contortus* strains (Molento and Prichard 1999). Nematodes have a whole battery of Pgp genes, which may be selected by the use of ivermectin (Prichard and Roulet 2007). More research is needed on these gene families to see whether they are present in parasitic species and whether they could play an important role in the efflux of MLs. In present state of knowledge, molecular epidemiology of MLs is still to be constructed.

11.5 The Threat of Resistance to Several Anthelmintic Dugs

The targets for anthelmintic resistance mechanism of nematodes are different for benzimidazoles, imidazothiazoles and macrocyclic lactones (Wolstenholm et al. 2004). Thus it is expected that resistance to one drug should not influence the resistance to another. Conversely, when multidrug resistance mechanisms (Kerboeuf et al. 2010)

are involved it would be expected that a common selection may occur in part, and that a certain degree of association of resistance may occur in field conditions. One may also suspect that farming practises that favour resistance to one drug may also influence the building up of resistance. The putative associations of benzimidazole and macrocyclic lactone resistances in the gastro-intestinal *Haemonchus contortus* (Mottier and Prichard 2008) or *Onchocerca volvulus* nematode in man (Eng et al. 2006) are also indicative for association between resistances. The large prevalence of resistance has favoured the marketing of drug associations in animals which may be harmful if resistances are co-selected or conversely protective if there is enhanced synergy between drugs metabolism (Alvarez et al. 2008). The developing use of associations of anthelmintic drugs to combat the increasing resistance to several drugs is a real bet since the exact mechanisms of resistance is not yet available (see the putative mechanisms of resistance against levamisole or macrocyclic lactones), the pharmacology of several drugs taken together may be unexpected. The molecular epidemiology of resistance to combination of several drugs is then completely obscure, unless much progress is done on mechanisms involved in each drug and their pharmacological behaviour in their combinations.

11.6 Conclusion

Anthelmintic control of helminths infections resulted in drug resistance appearance in many species. To study the factors associated to drug resistance selection, frequency of resistant parasites has to be monitored. Most parasites cannot be studied by direct methods (population size estimated by a capture-recapture analysis or migration studied by survey of movements of identified animals, for instance) and their biology has to be assessed via indirect means, most notably using molecular markers. This can be performed only in benzimidazole resistant isolates since molecular mechanisms of anthelmintic resistance are not yet fully known for LEV and MLs, and potential genetic markers are few. Moreover, anthelmintic resistance may be suspected by mistake: false resistance may arise from numerous situations, from generic drugs that may be counterfeit to co-administration of two drugs that reduce bioavailability of one drug (Cabaret 2010). More work is required in the field of molecular mechanisms conferring anthelmintic resistance before these leads can provide practical markers for anthelmintic resistance and allow a better understanding of the underlying forces selecting for resistance.

References

Alvarez L, Lifschitz A, Entrocasso C et al (2008) Evaluation of the interaction between ivermectin and albendazole following their combined use in lambs. J Vet Pharmacol Ther 31:230–239

Bartley DJ, Jackson F, Jackson E et al (2004) Characterisation of two triple resistant field isolates of *Teladorsagia* from Scottish lowland sheep farms. Vet Parasitol 123:189–199

Beech RN, Silvestre A (2010) Mutations associated with anthelmintic drug resistance. Anti-Infect Agents Med Chem 9:105–112

Beech RN, Prichard RK, Scott ME (1994) Genetic variability of the beta-tubulin genes in benzimidazole-susceptible and -resistant strains of *Haemonchus contortus*. Genetics 138:103–110

Besier B, Knox M, Love S (2003) Individual sheep management – new opportunities for objective parasite control. Wool Tech Sheep Breed 51:146–147

Blouin MS, Yowell CA, Courtney CH et al (1995) Host movement and the genetic structure of populations of parasitic nematodes. Genetics 141:1007–1014

Boulin T, Gielen M, Richmond JE et al (2008) Eight genes are required for functional reconstitution of the *Caenorhabditis elegans* levamisole-sensitive acetylcholine receptor. Proc Natl Acad Sci USA 105:18590–18595

Bourguinat C, Ardelli BF, Pion SDS et al (2008) P-glycoprotein-like protein, a possible genetic marker for ivermectin resistance selection in *Onchocerca volvulus*. Mol Biochem Parasitol 158:101–111

Cabaret J (2010) False resistance to antiparasitic drugs: causes from shelf availability to patient compliance. Anti-Infect Agents Med Chem 9:161–167

Coles GC (2002) Cattle nematodes resistant to anthelmintics: why so few cases? Vet Res 33:481–489

Cully DF, Paress PS (1991) Solubilization and characterization of a high affinity ivermectin binding site from *Caenorhabditis elegans*. Mol Pharmacol 40:326–332

Cully DF, Vassilatis DK, Liu KK et al (1994) Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. Nature 371:707–711

Dent JA, Smith MM, Vassilatis DK et al (2000) The genetics of ivermectin resistance in *Caenorhabditis elegans*. Proc Natl Acad Sci USA 97:2674–2679

Elard L (1998) La résistance aux benzimidazoles chez *Teladorsagia circumcincta*, nématode parasite de petits ruminants Etude du déterminisme génétique et recherche des conséquences sur la fitness des parasites Université François Rabelais, Tours

Elard L, Humbert JF (1999) Importance of the mutation of amino acid 200 of the isotype 1 beta-tubulin gene in the benzimidazole resistance of the small-ruminant parasite *Teladorsagia circumcincta*. Parasitol Res 85:452–456

Elard L, Comes AM, Humbert JF (1996) Sequences of beta-tubulin cDNA from benzimidazole-susceptible and -resistant strains of *Teladorsagia circumcincta*, a nematode parasite of small ruminants. Mol Biochem Parasitol 79:249–253

Elard L, Sauve C, Humbert JF (1998) Fitness of benzimidazole-resistant and -susceptible worms of *Teladorsagia circumcincta*, a nematode parasite of small ruminants. Parasitology 117:571–578

Eng JKL, Blackhall WJ, Osei-Atweneboana MY, Bourguinat C, Galazzo D, Beech RN, Unnasch TR, Awadzi K, Lubega GW, Prichard RK (2006) Ivermectin selection on betatubulin: evidence in Onchocerca volvulus and Haemonchus contortus. Mol Biochem Parasitol 150:229–235

Enyati A, Hemingway J (2010) Malaria management: past, present and future. Annu Rev Entomol 55:569–591

Fleming JT, Squire MD, Barnes TM et al (1997) *Caenorhabditis elegans* levamisole resistance genes lev-1, unc-29, and unc-38 encode functional nicotinic acetylcholine receptor subunits. J Neurosci 17:5843–5857

Geary TG, Sims SM, Thomas EM et al (1993) *Haemonchus contortus*: ivermectin-induced paralysis of the pharynx. Exp Parasitol 77:88–96

Gilleard JS, Beech RN (2007) Population genetics of anthelmintic resistance in parasitic nematodes. Parasitology 134:1133–1147

Grillo V, Jackson F, Gilleard JS (2006) Characterisation of *Teladorsagia circumcincta* microsatellites and their development as population genetic markers. Mol Biochem Parasitol 148:181–189

Hoekstra R, Criado-Fornelio A, Fakkeldij J et al (1997) Microsatellites of the parasitic nematode *Haemonchus contortus*: polymorphism and linkage with a direct repeat. Mol Biochem Parasitol 89:97–107

Hu Y, Xiao SH, Aroian RV (2009) The new anthelmintic tribendimidine is an L-type (levamisole and pyrantel) nicotinic acetylcholine receptor agonist. Plos Negl Trop Dis. doi:10.1371/journal.pntd.0000499

Huson DH (1998) SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14:68–73

Jones AK, Buckingham SD, Sattelle DB (2005) Chemistry-to-gene screens in *Caenorhabditis elegans*. Nat Rev Drug Discov 4:321–330

Kaminsky R, Ducray P, Jung M (2008) A new class of anthelmintics effective against drug-resistant nematodes. Nature 452:176–180

Kaplan RM (2004) Drug resistance in nematodes of veterinary importance: a status report. Trends Parasitol 20:477–481

Kerboeuf D, Guegnard F, Le Vern Y (2003) Detection of P-glycoprotein-mediated multidrug resistance against anthelmintics in *Haemonchus contortus* using anti-human mdr1 monoclonal antibodies. Parasitol Res 91:79–85

Kerboeuf D, Riou M, Neveu C et al (2010) Membrane drug transport in helminths. Anti-Infect Agents in Med Chem 9:113–129

Kopp SR, Coleman GT, Traub RJ et al (2009) Acetylcholine receptor subunit genes from *Ancylostoma caninum*: altered transcription patterns associated with pyrantel resistance. Int J Parasitol 39:435–441

Kwa MSG, Veenstra JG, Roos MH (1994) Benzimidazole resistance in *Haemonchus contortus* is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1. Mol Biochem Parasitol 63:299–303

Kwa MSG, Veenstra JG, Dijk MV et al (1995) Beta-tubulin genes from the parasitic nematode *Haemonchus contortus* modulate drug resistance in *Caenorhabditis elegans*. J Mol Biol 246:500–510

Lacey E, Prichard RK (1986) Interactions of benzimidazoles (BZ) with tubulin from BZ-sensitive and BZ-resistant isolates of *Haemonchus contortus*. Mol Biochem Parasitol 19:171–181

Lawrence KE, Rhodes AP, Jackson R et al (2006) Farm management practices associated with macrocyclic lactone resistance on sheep farms in New Zealand. N Z Vet J 54:283–288

Leignel V, Cabaret J (2001) Massive use of chemotherapy influences life traits of parasitic nematodes in domestic ruminants. Funct Ecol 15:569–574

Leignel V, Humbert JF (2001) Mitochondrial DNA variation in benzimidazole-resistant and -susceptible populations of the small ruminant parasite, *Teladorsagia circumcincta*. J Hered 92:503–506

Leignel V, Cabaret J, Humbert JF (2002) New molecular evidence that *Teladorsagia circumcincta* (Nematoda: Trichostrongylidea) is a species complex. J Parasitol 88:135–140

Leignel V, Silvestre A, Humbert JF, Cabaret J (2010) Alternation of anthelmintic treatments: a molecular evaluation for benzimidazole resistance in nematodes. Vet Parasitol 172:80–88

Martin RJ, Robertson AP (2007) Mode of action of levamisole and pyrantel, Anthelmintic Resistance, E153 and Q57. Parasitology 134:1093–1104

McCavera S, Rogers AT, Yates DM et al (2009) An ivermectin-sensitive glutamate-gated chloride channel from the parasitic nematode *Haemonchus contortus*. Mol Pharmacol 75:1347–1355

Melo AC (2005) Resistance to benzimidazoles in nematoda *Haemonchus contortus* in Ceara state. PhD thesis, Fortaleza University, Ceara, Brazil, 170 pp

Molento MB, Prichard RK (1999) Effects of the multidrug-resistance-reversing agents verapamil and CL 347,099 on the efficacy of ivermectin or moxidectin against unselected and drug-selected strains of *Haemonchus contortus* in jirds (*Meriones unguiculatus*). Parasitol Res 85:1007–1011

Mottier MD, Prichard RK (2008) Genetic analysis of a relationship between macrocyclic lactone and benzimidazole anthelmintic selection on *Haemonchus contortus*. Pharmacogenet Genom 18:129–140

Neveu C, Charvet C, Fauvin A et al (2007) Identification of levamisole resistance markers in the parasitic nematode *Haemonchus contortus* using a cDNA-AFLP approach. Parasitology 134:1105–1110

Neveu C, Charvet C, Fauvin A et al (2010) Genetic diversity of levamisole receptor subunits in parasitic nematode species and abbreviated transcripts associated with resistance. Pharmacogenet Genom 20:414–425

Njue AI, Prichard RK (2004) Genetic variability of glutamate-gated chloride channel genes in ivermectin-susceptible and -resistant strains of *Cooperia oncophora*. Parasitology 129:741–751

Otsen M, Plas ME, Lenstra JA et al (2000) Microsatellite diversity of isolates of the parasitic nematode *Haemonchus contortus*. Mol Biochem Parasitol 110:69–77

Palcy C, Silvestre A, Sauve C et al (2010) Benzimidazole resistance in *Trichostrongylus axei* in sheep: long-term monitoring of affected sheep and genotypic evaluation of the parasite. Vet J 183:68–74

Pion SDS, Nana-Djeunga H, Bourguinat C et al (2009) Dynamics of *Onchocerca volvulus* microfilarial loads of cameroonian patients submitted to repeated (5–23) ivermectin treatments over 14 years (1994–2007). Am J Trop Med Hyg 79:339

Pomroy WE (2006) Anthelmintic resistance in New Zealand: a perspective on recent findings and options for the future. N Z Vet J 54:265–270

Prichard RK, Roulet A (2007) ABC transporters and beta-tubulin in macrocyclic lactone resistance: prospects for marker development. Parasitology 134:1123–1132

Prichard RK, Oxberry M, Bounhas Y et al (2000) Polymerisation and benzimidazole binding assays with recombinant a- and b-tubulins from *Haemonchus contortus* In: American association of veterinary parasitologists, 45th annual meeting, Salt Lake City Marriott, p 81

Riley LW (2004) Molecular epidemiology of infectious diseases: principles and practices. ASM Press, Washington, DC, 1752 N st NW

Robertson SJ, Martin RJ (1993) Levamisole-activated single-channel currents from muscle of the nematode parasite *Ascaris-suum*. Br J Pharmacol 108:170–178

Robertson AP, Bjorn HE, Martin RJ (1999) Resistance to levamisole resolved at the single-channel level. FASEB J 13:749–760

Rufener L, Kaminsky R, Maser P (2009) In vitro selection of *Haemonchus contortus* for benzimidazole resistance reveals a mutation at amino acid 198 of beta-tubulin. Mol Biochem Parasitol 168:120–122

Sangster NC, Riley FL, Wiley LJ (1998) Binding of [3 H]m-aminolevamisole to receptors in levamisole-susceptible and -resistant *Haemonchus contortus*. Int J Parasitol 28:707–717

Sargison ND, Jackson F, Bartley DJ et al (2007) Observations on the emergence of multiple anthelmintic resistance in sheep flocks in the south-east of Scotland. Vet Parasitol 145:65–76

Shayan P, Eslami A, Borji H (2007) Innovative restriction site created PCR-RFLP for detection of benzimidazole resistance in *Teladorsagia circumcincta*. Parasitol Res 100:1063–1068

Silvestre A, Cabaret J (2002) Mutation in position 167 of isotype 1 beta-tubulin gene of Trichostrongylid nematodes: role in benzimidazole resistance. Mol Biochem Parasitol 120:297–300

Silvestre A, Humbert JF (2000) A molecular tool for species identification and benzimidazole resistance diagnosis in larval communities of small ruminant parasites. Exp Parasitol 95:271–276

Silvestre A, Humbert JF (2002) Diversity of benzimidazole-resistance alleles in populations of small ruminant parasites. Int J Parasitol 32:921–928

Silvestre A, Leignel V, Berrag B et al (2002) Sheep and goat nematode resistance to anthelmintics: pro and cons among breeding management factors. Vet Res 33:465–480

Silvestre A, Sauve C, Cortet J et al (2009) Contrasting genetic structures of two parasitic nematodes, determined on the basis of neutral microsatellite markers and selected anthelmintic resistance markers. Mol Ecol 18:5086–5100

Taly A, Corringer PJ, Guedin D et al (2009) Nicotinic receptors: allosteric transitions and therapeutic targets in the nervous system. Nat Rev Drug Discov 8:733–750

Tiwari J, Kumar S, Kolte AP et al (2006) Detection of benzimidazole resistance in *Haemonchus contortus* using RFLP-PCR technique. Vet Parasitol 138:301–307

Towers PR, Edwards B, Richmond JE (2005) The *Caenorhabditis elegans* lev-8 gene encodes a novel type of nicotinic acetylcholine receptor alpha subunit. J Neurochem 93:1–9

Vermunt JJ, West DM, Pomroy WE (1995) Multiple resistance to ivermectin and oxfendazole in Cooperia species of cattle in New Zealand. Vet Rec 137:43–45

von Samson-Himmelstjerna G, Walsh TK, Donnan AA et al (2009) Molecular detection of benzimidazole resistance in *Haemonchus contortus* using real-time PCR and pyrosequencing. Parasitology 136:349–358

Walsh TK, Donnan AA, Jackson F et al (2007) Detection and measurement of benzimidazole resistance alleles in *Haemonchus contortus* using real-time PCR with locked nucleic acid Taqman probes. Vet Parasitol 144:304–312

William S, Sabra A, Ramzy F et al (2001) Stability and reproductive fitness of *Schistosoma mansoni* isolates with decreased sensitivity to praziquantel. Int J Parasitol 31:1093–1100

Williamson SM, Robertson AP, Brown L et al (2009) The nicotinic acetylcholine receptors of the parasitic nematode *Ascaris suum*: formation of two distinct drug targets by varying the relative expression levels of two subunits. PLoS Pathog. doi:10.1371/journal.ppat.1000517

Wolstenholme AJ, Fairweather I, Prichard R, von Samson-Himmelstjerna G, Sangster NC (2004) Drug resistance in veterinary helminths. Trends Parasitol 20:469–476

Xu M, Molento M, Blackhall W et al (1998) Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. Mol Biochem Parasitol 91:327–335

Chapter 12
Molecular Epidemiology of Disease Resistance Genes with Perspectives for Researches on Biological Invasions and Hybrid Zones

Nathalie Charbonnel and Jean-François Cosson

Abstract Molecular epidemiology has largely been investigated from the parasite side, therefore contributing to a better understanding of transmission patterns and factors influencing disease spread. From the hosts' point of view, molecular epidemiology has lead to the identification of components of susceptibility and resistance to infectious diseases. In this chapter, we aim to explain why resistance to infections exhibits such a remarkable degree of polymorphism while being resistant obviously confers a high selective advantage to hosts. In this context, we develop host molecular epidemiology with regard to the concepts of evolutionary biology and immunoecology. We detail the mechanisms that are likely to underlie the variable degrees of host resistance polymorphism observed among natural populations. Particular attention is given to recently emphasized topics, including the risks of immunopathology, the spatial structure of populations, the impact of neutral evolutionary processes and the phenotypic plasticity of resistance. We evidence the consequences of this polymorphism for disease epidemiology both from empirical examples and genetic epidemiological modelling of resistance evolution. We stress the numerous gaps that remain to be explored to understand these patterns of disease resistance polymorphism. We particularly emphasize the cruel lack of theoretical predictions and data that have focused on biological invasions and hybridisation, two biological situations of main interest for emergence.

N. Charbonnel (✉) • J.-F. Cosson
INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus international de Baillarguet, CS 30016, F-34988 Montferrier-sur-Lez cedex, France
e-mail: nathalie.charbonnel@supagro.inra.fr; cosson@supagro.inra.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_12,

12.1 Introduction

Infectious diseases represent an increasing and substantial threat to global health (Daszak et al. 2000; King et al. 2006; Morens et al. 2004). In humans, they account for 25% of mortality and a similar fraction of morbidity (WHO 2005; Morens et al. 2004). The damages they impose to livestock and crops cost several billions of euros each year ($41 billions per year in the USA, Daszak et al. 2000; King et al. 2006). Several socioeconomic and environmental factor risks of infectious disease emergence have now been statistically identified (Jones et al. 2008). Among them, anthropogenic global climate changes, which affect the distribution and life cycles of organisms, as well as travel, migration, and trade, which promote the spread of infections into new populations, are of main importance (Daszak et al. 2000; King et al. 2006; Morens et al. 2004). Both stress the relevance of pathogen/parasite transmission and spatialisation in the epidemiological risks of emerging infectious diseases (Deter et al. 2010; Ostfeld et al. 2005).

Recent advances in molecular epidemiology have largely contributed to improve inferences of transmission and environmental factors influencing disease spread (review in Archie et al. 2009). The combination of these approaches with evolutionary biology now provides insights into the neutral and adaptive processes shaping the spatio-temporal distribution of infectious organisms and governing the evolution of virulence (reviewed in Archie et al. 2009; Galvani 2003; Grenfell et al. 2004). These aspects are largely investigated in several chapters of this book.

From the hosts' point of view, genetic epidemiological approaches have been conducted to identify components of predisposition to infectious diseases. These researches, which are based on association studies and genome-wide screens, still mainly concern humans and models of veterinary importance (reviews in Charbonnel et al. 2006; Cooke and Hill 2001; Quintana-Murci et al. 2007). They have evidenced that a considerable fraction of the genome is dedicated to defence against infectious organisms (5% of genes in most mammals, Trowsdale and Parham 2004). Several major gene loci have been mapped with regard to susceptibility to certain diseases. Molecular and evolutionary biologists are now working on the detection of geographical variants, on the links between genetic variations and encoded protein function or expression, and on the detection of signatures of selection shaping the spatio-temporal genetic differentiation observed among populations (Bradbury 2004). Studies investigating all these axes of researches have preferentially focused on Major Histocompatibility Complex (*Mhc* or *Hla* in humans) genes (Acevedo-Whitehouse and Cunningham 2006), but recent immunogenomic approaches have allowed interesting developments of non-*Mhc* gene investigations (Barreiro and Quintana-Murci 2010).

In natural populations, researches have lead to paradoxical and intriguing patterns: Being resistant to diseases confers a high selective advantage to hosts, yet many host-pathogen systems show a remarkable degree of polymorphism of host resistance (Jarosz 1991; Thrall and Burdon 2003). For example, alleles conferring mannose-binding lectin (MBL) deficiency are very common worldwide although such insufficiency is associated with increased susceptibility to several infectious

diseases in humans (Eisen and Minchinton 2003). Besides, in agronomy, identifications and further manipulations of host resistance genetics have received particular attention for management and enhancement programs (Restif 2009). The possibilities of developing strains that have increased resistance to infectious diseases are explored, but epidemio-genetic models are required for decision making process as they will further help identify the relative impacts of different resistance genes on the dynamics of infections (Nath et al. 2004).

In this chapter, we propose to explore potential responses to both aspects: the polymorphism observed at resistance genes in natural populations and the consequences of this polymorphism for disease epidemiology. We will mainly, but not exclusively, focus on zoonotic infectious diseases and evolutionary biology, as they are our main domains of expertise. We will first precise our definition of resistance, as many confounding designations may be found in the literature. We will review the evolutionary processes shaping and maintaining polymorphism at host resistance genes, and we will detail the consequences of these spatio-temporal genetic variations in resistance on the epidemiology and risks of infectious disease emergence. Lastly, we will focus on two biological situations that we think could be of high risks for infectious disease emergence: biological invasions and hybrid zones. These situations are indeed of high relevance to explore the potential roles of resistance genes in facilitating/preventing the emergence of infectious diseases in new host species and/or new environments.

12.2 Resistance and Its Genetic Determinisms

For convenience, we will use ‘parasite’ throughout this chapter as a generic terminology for both micro- and macroparasites.

12.2.1 Definitions

Resistance in host – parasite interactions is the host ability to limit parasite burden during an infection (Svensson and Raberg 2010). It is generally measured as the proportion of uninfected individuals or as the inverse of parasite load (Raberg et al. 2009; Restif and Koella 2004).

Avoidance (including behavioral strategies and barriers to infection) as well as processes that limit the spread of infection or rapidly clear it within the host (e.g. immune responses) are considered as forms of resistance (Miller et al. 2005; Restif and Koella 2004). They all inhibit the spread of infection by reducing the survival and reproductive potential of the parasite (Roy and Kirchner 2000). Resistance has hence a direct negative effect on parasite fitness (Rausher 2001). We would like here to emphasize the important differences existing between resistance and another defence mechanism, tolerance, which is the ability to reduce the negative impacts of infection on host fitness. These two traits are sometimes used synonymously

although their epidemiological and evolutionary implications are highly different, as pointed out by several theoreticians (e.g. Raberg et al. 2009; Restif and Koella 2004; Roy and Kirchner 2000). In particular, tolerance does not inhibit the growth or reproduction of parasites (Rausher 2001), and thus does not impose any direct cost on the parasite (Raberg et al. 2009; Svensson and Raberg 2010). In turn, there should not be any selection on the parasite to overcome this type of defence (Raberg et al. 2009). As resistance and tolerance might sometimes be genetically correlated (Raberg et al. 2007; Simms and Triplett 1997), resulting trade-offs between these types of defences will further affect the prevalence of parasites in natural populations, the epidemiology of infectious diseases, and in turn, the evolution of resistance.

12.2.2 Genetic Determinisms of Resistance to Parasites

Genetic (heritable) and environmentally induced variation in resistance have been found in plants and animals (review in Raberg et al. 2009).

12.2.2.1 Genetic Variation (G)

Many candidate and QTL gene studies have allowed to identify susceptibility/resistance alleles to diseases, therefore providing evidence that these genes and their biological pathways are relevant to specific diseases. Such approaches have been successfully applied in humans (e.g. Bradbury 2004; Cooke and Hill 2001), domestic (e.g. Bishop et al. 2002), laboratory (e.g. for mouse resistance to Orthopoxviruses: Stanford et al. 2007) and wild (e.g. for bovin resistance to Trypanosoma Gautier et al. 2009) organisms. They concern all forms of resistance (e.g. entry into host cells, limitation of proliferation, see Nath et al. 2004). It is likely that the tremendous advances in high throughput sequencing technologies will further facilitate the identification of genetic variations that are involved in the risks of infectious diseases, especially for non-laboratory organisms.

Several mathematical models of genetic interactions have been developed to understand host-parasite coevolution. They include locus-based, quantitative and adaptive dynamics resistance models. They may lead to major differences in host-parasite coevolution outcomes. For example, a qualitative form of resistance is expected to reduce parasite virulence, while a quantitative form will generally increase it (Gandon and Michalakis 2000). Briefly, the two major locus-based models include ‘gene-for-gene’ (GFG, Flor 1956), with one parasite genotype having ‘universal virulence’ and being able to infect all host genotypes, and ‘matching alleles’ (MA, Frank 1994), with resistance requiring a specific match between the host and its parasite (see review in Agrawal and Lively 2002). Although GFG mostly concerns plant-parasite interactions, it has also been proven to be relevant for innate genes of animals (e.g. in the Drosophila melanogaster and sigma virus, Bangham

et al. 2008). These GFG and MA models, as well as all in-between ones, can only be extended to a small number of genes (Sasaki 2000). But resistance to parasites is often a polygenic trait, especially in vertebrates (Woolhouse et al. 2002). As such, quantitative genetic models have been developed to fit with these complex genetic interactions. They assume that a large number of loci make a small additive contribution to resistance (Boots et al. 2009). Different methods have been used to explore the evolution of resistance phenotypic quantities, but we will not detail them here (Abrams 2001). Only note that these quantitative models do not explicit any mutation process. Mutations arise implicitly due to a probability distribution, therefore maintaining many genetic variations at low density (Boots et al. 2009). By contrast, the 'adaptive dynamics' approach, based on the game theory, explicitly describes how mutations arise, thus restricting the set of mutations that occur (Boots et al. 2009). This modelling examines the invasion of new mutants into a monomorphic resident population.

12.2.2.2 Phenotypic Plasticity (GxE)

Another component that may influence the evolution of resistance gene is phenotypic plasticity, i.e. the interactions between genetics and environment (GxE). It refers to the ability of a genotype to exhibit alternative phenotypes in response to environmental conditions (Garland and Kelly 2006). Lingappa et al. (2004) documented the first GxE interaction affecting susceptibility to an infectious disease. They showed that the risk to leptospirosis infections among triathletes during a lake swim was linked to the interaction between the haplotype *Hla-Dq6* and the swallowing of water. Different softwares have since been developed to detect these GxE interactions using case control studies (Amato et al. 2010). They have allowed the identification of several GxE interactions influencing risks of human diseases (see a review in Hunter 2005). In natural animal populations, the most obvious examples concern the variations of disease resistance between casts of social insects. Although these casts exhibit the same genetic background, they may be associated with different environmental conditions during their life, which will further affect the expression of resistance genes (e.g. in ants: Bocher et al. 2007). Two other facets of phenotypic plasticity can be emphasized. The first one concerns the potential interactions between multiple infections. Resistance to a given parasite may depend on host gene expression, which can itself be modulated by other parasite genome fractions, such as immune modulatory genes (e.g. Coscoy 2007) or micro-RNA (refs in Li et al. 2010). Either enhancement or decrease in resistance levels might be expected during such co-infections (see examples in Woolhouse et al. 2002). The second one refers to immune memory in vertebrates. It can be interpreted as phenotypic plasticity, as secondary responses will be greater than primary ones without any changes in the host genetic background (Sadd and Schmid-Hempel 2009).

Although phenotypic plasticity has been recognized as a central concept in the study of complex diseases (Amato et al. 2010), there are still few studies investigating potential GxE interactions. It is however of main importance both for a better

estimation of risk disease associated with genetic variants and for evolutionary epidemiological modelling. Indeed, GxE interactions will reduce the importance of the host genotype–parasite genotype interactions that drive co-evolution. They will also allow to include explicitly the potential role of ecological feedbacks on the evolution of resistance (Best et al. 2008; Boots et al. 2009). For example, such phenotypic plasticity could give support to the hypothesis of exposure risks mediating changes in genetic resistance: as resistance alleles spread within a host population, parasite prevalence could decrease, and the selection for resistance could in turn decrease (e.g. in locusts (Wilson et al. 2002) and in ants (Bocher et al. 2007)).

12.3 Neutral and Selective Evolutionary Forces Interact to Shape Patterns of Host Resistance Gene Polymorphism over Different Timescales

While biomedical or breeding sciences have long focused on the associations between genetic variants and infectious diseases, evolutionary biology has recently 'infected' immunology, with the main aim of explaining how genetic variation is maintained in wild populations. It explores the relative influence of four main evolutionary processes, i.e. mutation, drift, gene flow and selection in shaping polymorphism and genetic differentiation observed at immune genes within and among populations. Whatever the historical and geographical scales considered, empirical studies have revealed the strong interplay between on one side mutation, drift, migration, and on the other side, selective pressures that act on host resistance genes. We give here some contrasted examples illustrating the diversity of outcomes that have been observed *in natura* as the result of different combinations of these evolutionary forces.

12.3.1 Phylogenetic Variations of Resistance Genes

Phylogeny of resistance genes may first reflect the global phylogeny of species. Pagès et al. (Pagès et al. Submitted) recently investigated the phylogenetic variations of β_3 chain integrin sequences among South-East Asian rodent species. They proposed from immunological studies (Gavrilovskaya et al. 1998, 1999; Geimonen et al. 2002; Mou et al. 2006) that genetic variation within the Plexin-Semaphorin-Integrin domain of β_3 chain integrin could reflect the possibility for a rodent species to carry a hantavirus, and could help to determine *a priori* whether this hantavirus is likely to be pathogenic or not for humans. They observed that the β_3 chain integrin genetic variation reflected the neutral phylogeny of the 15 rodent species studied. Besides, they found two nucleotide sites that seemed informative to discriminate between rodents that are reservoirs of human pathogenic hantaviruses and rodents that are non reservoir of hantaviruses. Only one of these sites seemed to evolve

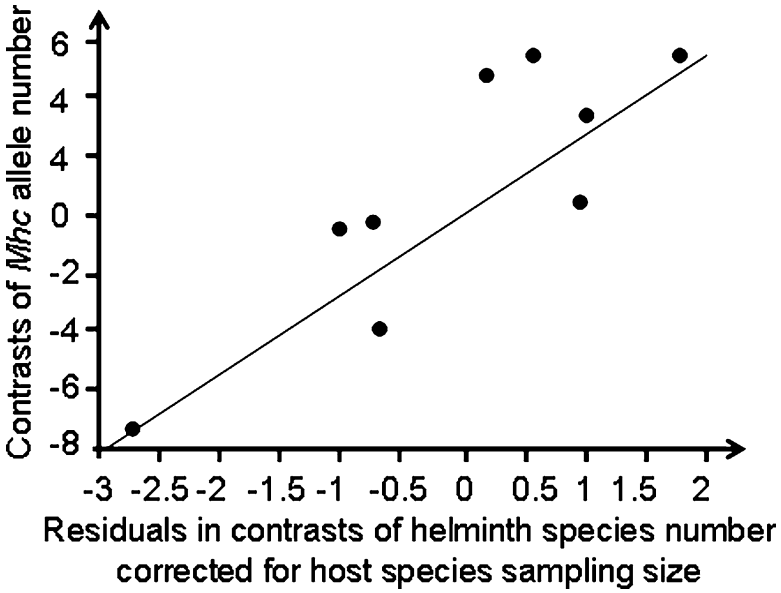

Fig. 12.1 Relationships between the contrasts in number of *Mhc* class II alleles and the residuals of contrasts in helminth species richness (controlled for host sample size). Data concern the *Dqa* and *Drb* gene variability assessed in ten rodent species (See details in Gouy de Bellocq et al. 2008)

under the action of positive selection, but only within rodent species that carried human pathogenic hantaviruses. Further investigations should confirm whether amino-acid changes at this position could mediate conformational changes that would prevent the fixation of pathogenic hantaviruses.

By contrast, Ferguson et al. (2008) revealed high levels of polymorphism at a gene encoding an oligoadenylate synthetase (Oas1b) in the house mouse *Mus musculus* and related species. This gene encodes for proteins that provide specific resistance against West Nile Virus infections. This variability was mediated by the long-term action of balancing selection. As for *Mhc* genes in mammals (Bryja et al. 2005; Edwards et al. 1997; Figueroa et al. 1988), for the *Trim5* gene in Old World monkeys (Newman et al. 2006) and for plant resistance genes (Bergelson et al. 2001), this polymorphism was very ancient (2.8 Myr) and predated speciation events. In all these cases, the long-term balancing selection resulted in trans-specific polymorphisms (Figueroa et al. 1988).

Finally, comparative analyses have evidenced that the interplay between species history and parasite-mediated selection could mediate host resistance gene polymorphism. Gouy de Bellocq et al. (2008) showed a positive relationship between the helminth diversity and the allelic diversity at two *Mhc* class II genes in rodent species (Fig. 12.1). This pattern was still significant when controlling for the (phylogenetic) non-independence of rodent species using the independent contrast method. Moreover, Gouy de Bellocq et al. (2008) also showed that this relationship

was not driven by neutral allelic diversity, *i.e.* effective population size of the different species, thus reinforcing the hypothesis of a direct selective pressure exerted by parasite richness on *Mhc* gene polymrophism.

12.3.2 Phylogeographic Variations of Resistance Genes

Within species, neutral historical forces as well as parasite-mediated selection participate in shaping resistance gene polymorphism. Prugnolle et al. (2005) indeed showed that while human colonisation history had been important in shaping the present patterns observed at human *Hla* genes, the diversity of viruses had also been essential in driving and maintaining the genetic diversity at these genes. Barreiro et al. (2005) emphasized that history and selection could lead to different outcomes of resistance gene polymorphism, even for genes sharing very close physical vicinity. Considering the genetic variation of two innate genes *Cd209* and *Cd209l* in human populations, they found traces of ancient population structure in Africa, and showed that the diversity of *Cd209* gene was strongly subjected to functional constraints while balancing selection was likely to generate the high levels of diversity observed at the *Cd209l* gene.

Phylogeographical analyses of wild vertebrate immune genes are still very scarce and none of them was applied in the context of zoonoses. Among the seven studies found in the literature, two revealed that *Mhc* gene polymorphism showed clear marks of the phylogeographical history of the species studied (Berggren et al. 2005; Langefors 2005), while four others evidenced selection acting on *Mhc* genes (e.g. Koutsogiannouli et al. 2009; Male et al. In revision), probably mediated by variations of parasite communities (e.g. Alcaide et al. 2008). A recent study investigated Tumor Necrosis Factor-alpha (*Tnf-*α) gene expression and promoter polymorphism among *Myodes glareolus* populations over Europe. This rodent is the natural reservoir of *Puumala virus* (Bunyaviridae, Hantavirus), the agent of Nephropatia Epidemica in humans. The phylogeographical pattern observed provided arguments in favor of selection mediating European spatial variations in *Tnf-*α gene expression (Fig. 12.2). Guivier et al. (2010) also showed that this polymorphism was strongly associated with both promoter sequence variations, what corroborates its genetic determinism, and PUUV endemicity over Europe. As TNF-α induces the antagonist effects of limiting PUUV infection and inducing inflammatory disorders, Guivier et al. (2010) proposed that polymorphism in *Tnf-*α gene expression could mediate a balance of tolerance/resistance to PUUV infections.

12.3.3 Among Population Variations of Resistance Genes

Population genetics has been widely applied to the study of evolutionary forces mediating polymorphism at immune genes. Contrasted patterns observed at immune vs neutral genes were considered as evidence of selection. A recent review clearly

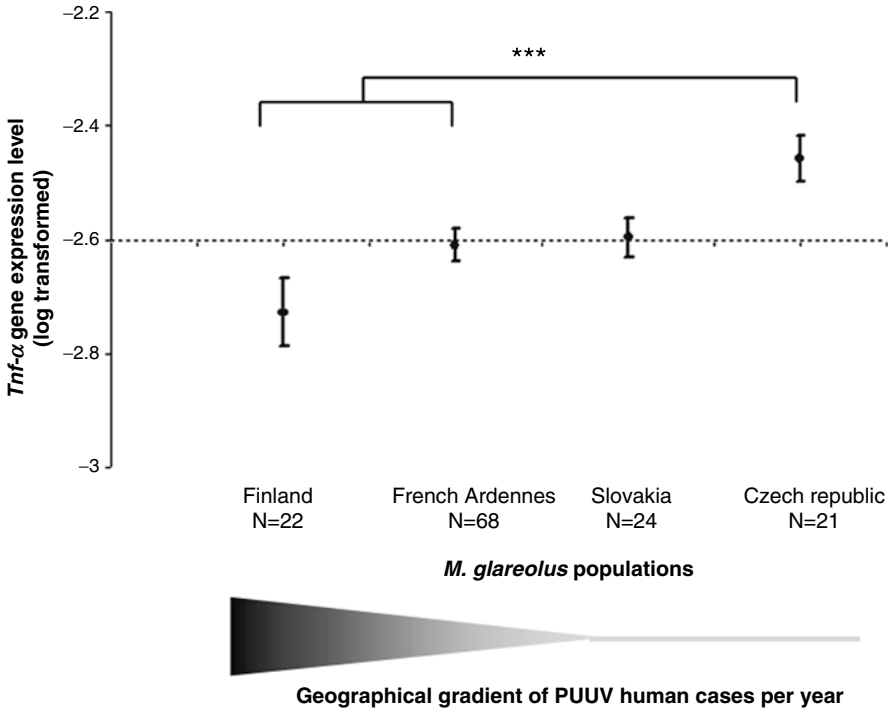

Fig. 12.2 Differences of *Tnf-α* gene expression between European populations of the bank vole *Myodes glareolus*. *** indicates pairs of populations that exhibit significantly different levels of *Tnf-α* gene expression (Post-hoc Tukey-Kramer test, $p<0.05$) (See details in Guivier et al. 2010)

described the different patterns expected under various modes of selection (Spurgin and Richardson 2010).

Many studies have not detected any contemporary selection at immune genes, therefore stressing the major role of genetic drift and demographic factors in shaping the variation and distribution of *Mhc* polymorphism (e.g. Boyce et al. 1997; Seddon and Ellegren 2004). By contrast, researches conducted on the montane water vole (*Arvicola scherman*, ex *terrestris*) have well illustrated how population genetics, demography, life history traits and parasitology may help to understand how immune genes evolve at the contemporary scale. This rodent species exhibits pluri-annual demographic cycles, with density varying from less than one to several hundred individuals per hectare over 5–6 years (Giraudoux et al. 1997; Saucy 1994) in the East of France. Berthier et al. (2006) emphasized that drift operating during low density populations (small size and geographical isolation of demes) and migration occurring principally while population size increased closely interacted to maintain high genetic diversity at neutral microsatellites. Besides, Bryja et al. (2007) investigated the relative influence of contemporary selection between these low and outbreak phases of density cycles on two *Mhc* class II gene polymorphism. Interestingly, they showed evidence of spatial and temporal variations in the interplay between

drift, immigration and selection acting on *Mhc* genes in the montane water vole. At low densities, significantly higher levels of genetic differentiation between demes were detected at *Mhc* genes than at microsatellites, suggesting the action of local selection at these immune genes. With increasing density, the spatial genetic structuring of both types of markers decreased. Nevertheless, once high levels of gene flow between demes were re-established, the *Dqa Mhc* gene exhibited significantly lower genetic differentiation among populations than microsatellites, what revealed the action of balancing selection acting on this *Mhc* gene (Fig. 12.3a). Further studies focusing on the helminth and virus communities (Tollenaere et al. 2008) of *A. scherman* in these populations have indicated the influence of parasite load and nematode species (*Trichuris arvicolae*) on the presence/absence of particular *Mhc* alleles (Fig. 12.3b). In addition, Charbonnel et al. (2010) revealed associations between *Mhc* genotypes and cellular-mediated immunity within these *A. scherman* populations (Fig. 12.4), as well as a trade off between this immune response and the development of a secondary sexual character (lateral scent glands). The immunocompetence handicap (ICHH Folstad and Karter 1992; Wedekind et al. 2006) could be another mechanism mediating these associations between *Mhc* polymorphism and the montane water vole life history traits.

Such applications of population genetics to the study of resistance gene polymorphism are critical for the prevention of infectious disease emergence. An increasing number of studies indicated that host genetic diversity plays an important role in buffering populations against parasites and widespread epidemics (Sommer 2005). Global changes leading to fragmentation, decline of population size or increase of inbreeding may contribute to diminish host resistance genetic diversity, what usually result in the rise of susceptibility to parasites and risks of disease emergence (e.g. Coltman et al. 1999; Sommer 2005).

12.4 What Prevents the Fixation of Resistance?

Polymorphism at resistance genes seems to be a very common pattern *in natura*, and determining the forces that prevent the fixation of resistance and shape the variability observed at resistance genes has been a major topics of evolutionary biology, involving both the development of theoretical modelling as well as empirical analyses.

12.4.1 Host-Parasite Coevolution

The examples detailed above suggest that host-parasite coevolutionary processes, *i.e.* the reciprocal, adaptive genetic changes occurring between these interacting species (Woolhouse et al. 2002), might underly the genetic polymorphism observed at resistance genes. Theoretical modelling has shown that many outcomes could

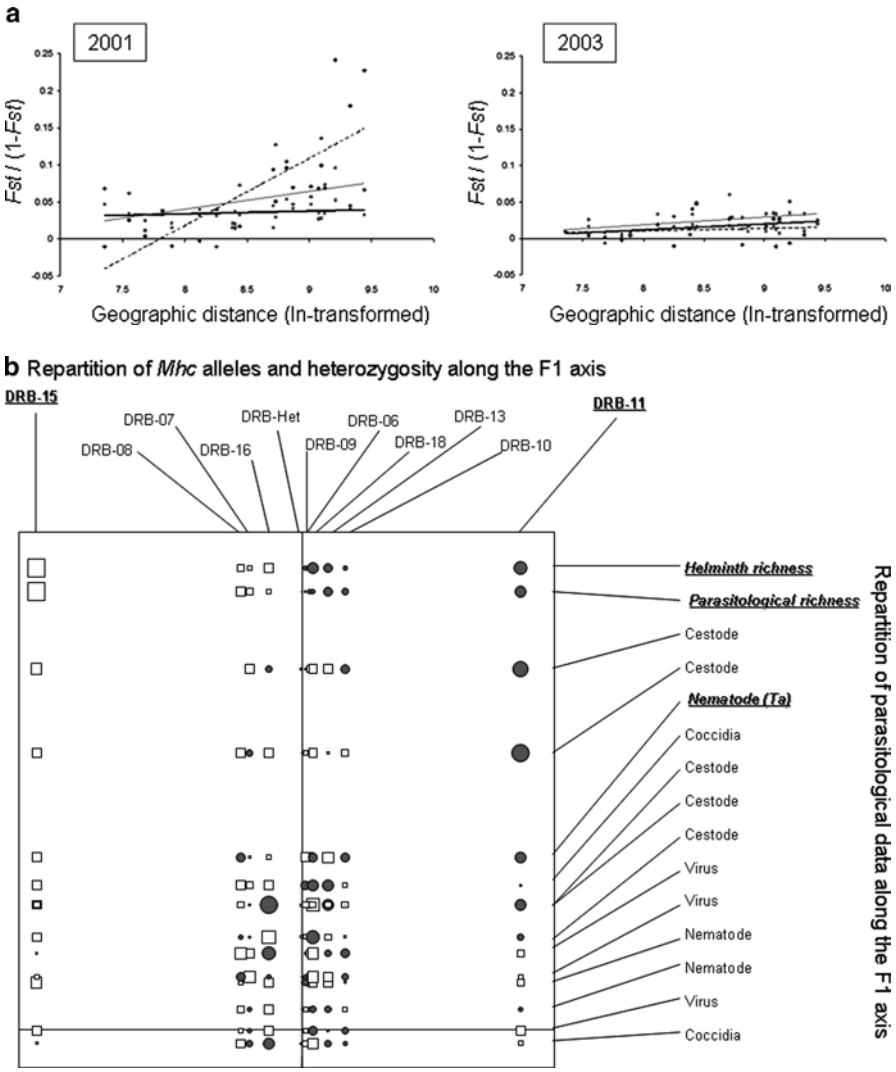

Fig. 12.3 *Mhc* class II polymorphism in *Arvicola scherman* populations exhibiting pluri-annual density cycles. (**a**) Correlation between genetic $F_{ST}/(1 - F_{ST})$ and geographical distance (ln-transformed) for two sampling dates corresponding to low density phase (2001) and outbreak (2003). These correlations are represented for 14 microsatellites (*stars* and *thick black line*) and for two *Mhc* class II genes, *Dqa* (*black diamond* and *dashed black line*), *Drb* (*grey diamond* and *grey solid line*). (**b**) Associations between DRB and parasitological variables projected on the F1 axis (31% of the total variance) of the co-inertia analysis between immunogenetic and parasitological data. *Grey circles* indicate positive associations and *white squares*, negative associations. The size of the symbol indicates the strength of the association. Associations confirmed by cross-validation are underlined. *Ta* refers to the nematode *Trichuris arvicolae* (See details in Bryja et al. 2005)

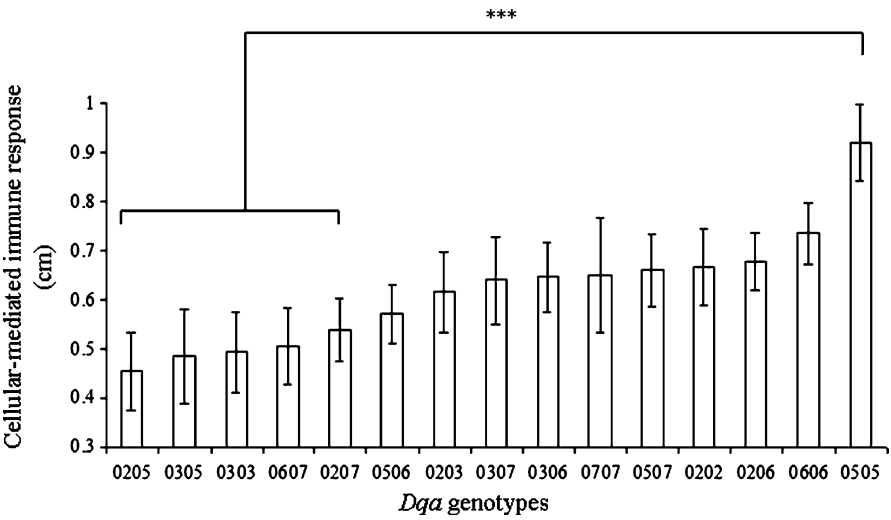

Fig. 12.4 Relationships between *Mhc* class II polymorphism and cellular-mediated immunity (response to phytohaemagglutinin)in *Arvicola scherman* populations. Error bars represent ±1 S.E. of the mean. *** indicate the heterozygous *Dqa* genotypes exhibiting significantly lower immunity than the homozygous *Dqa* 0505 genotype (Post-hoc Tukey-Kramer tests, $p<0.05$) (See details in Charbonnel et al. 2010)

arise from coevolution, including stable polymorphisms ('trench warfare' hypothesis Stahl et al. 1999), dynamic polymorphisms with cyclic or chaotic fluctuations in allele frequencies of varying amplitude ('Red Queen' dynamics, Hamilton et al. 1990) or transient polymorphism with recurrent selective sweeps of different favorable alleles, appearing through mutation or migration, and becoming fixed in the population ('arms race' model, Holub 2001).

The actions of both positive selection, which makes many resistance genes evolve rapidly at the protein level, and balancing selection, which can result from heterozygote advantage, frequency-dependent selection or selection that varies in time and space (see a review in Landry and Bernatchez 2001), contribute to maintain resistance gene polymorphism in populations. In addition to the examples provided above, the observation that epidemics could favour resistant host genotypes (Duffy and Sivars-Becker 2007; Duncan and Little 2007) provided evidence that these adaptive coevolutionary processes acting on resistance genes could be mediated by parasites. Similarly, the existence of coincident clines in infection rates and genetic disease resistance, as documented by Springer (2007) in a serpentine flax-flax rust interaction or by Guivier et al. (2011) in the *M. glareolus – Puumala virus* system in the French Ardennes, corroborated these strong links between parasite-mediated selection and host resistance polymorphism.

The genetic bases underlying host-parasite coevolutionary processes will strongly interfere with the outcomes described above. In particular, the specificity of resistance, which is inherent to the MA model, is expected to give rise to adaptation to

the most common parasite genotype. This will result in a negative indirect frequency dependent advantage for rare genotypes, and consequently in dynamic polymorphism (e.g. Gandon et al. 1996; Gandon and Michalakis 2002). A necessary condition for stable long term polymorphism to occur is direct frequency-dependent selection (Tellier and Brown 2007). The rate of selection for the resistant allele should decrease with the increase of its own frequency. Such condition is fulfilled for diseases involving parasites that undergo several generations per host generation (Tellier and Brown 2007). By contrast, the GFG model is more problematic with regard to the maintenance of host resistance polymorphism. Indeed, it predicts that 'super-races' of parasites, i.e. those carrying virulence effective against all resistant genes present in a population, should evolve. This would result in a neutral evolution of resistance. Different selective and demographic mechanisms have further been identified as critical factors preventing the fixation of favorable host resistance genetic variants. They included the potential costs of resistance, which trades off with other fitness components, the spatial organization of host populations and the stochasticity that is inherent to the finite size of populations.

12.4.2 Costs of Resistance

12.4.2.1 Evolutionary Costs

The most commonly proposed explanation of GFG resistance polymorphism is that resistance genes are costly and that there is selection against those genes when they are unnecessary, i.e. in the absence of infection. The negative feedbacks between the prevalence of resistant hosts and their fitness advantage will impede the spread of resistance genes in host populations, and result in resistance gene polymorphism (Best et al. 2008). Indeed, at some point, the risks of infection will be so low that the benefits of resistance will no longer outweigh its costs, and the resistance trait will stop spreading. From a risk management point of view, this has important consequences as it implies that diseases cannot be eliminated by natural or breeding selection for host resistance traits (Roy and Kirchner 2000).

A common assumption is that these costs arise due to antagonistic pleiotropy, where the allele encoding for resistance has other detrimental effects on fitness in the absence of disease (Brown 2003). This mechanism underlies evolutionary costs of resistance and evolutionary trade-offs between resistance and other life history traits (e.g. between fecundity and bacterial resistance in Drosophila sp., McKean et al. 2008). Interestingly, trade-offs between different components of the immune defence system might also occur and underly evolutionary costs for particular resistance genes (Cotter et al. 2004; Schmid-Hempel 2003). A particular form of trade offs between different resistance traits could concern specific responses to given parasite strains/species. The specificity of some host-parasite interactions (e.g. Carius et al. 2001; Sadd and Schmid-Hempel 2009; Schmid-Hempel 2001) could lead to the persistence of dynamic polymorphisms due to selective fluctuations

in allelic frequency. Negative frequency-dependence driven by coevolving parasites would then maintain genetic variability for resistance ('Red Queen' model, see above). Besides, if parasite abundance varies between years, then fluctuating costs of resistance might generate temporal variation in selection which, under certain conditions, could maintain polymorphism at resistance genes within a population (Hedrick 1974, 1976).

The existence of such costs is supported by both theoretical arguments (Stearns 1992) and empirical evidence in insects (e.g. Moret and Schmid-Hempel 2000), birds (e.g. Nordling et al. 1998) and mammals (e.g. Devevey and Christe 2009). These costs have also been evidenced in diverse fitness traits including reduced survivorship (e.g. Devevey and Christe 2009), reduced fertility (Carius et al. 2001) or reduced competitive ability (Kraaijeveld and Godfray 1997). The significance of these evolutionary costs to the host is often not investigated and different studies have failed to demonstrate any (e.g. Simms and Triplett 1994; Williams et al. 1999). These results have therefore stimulated a large amount of debate about the ubiquity of evolutionary resistance costs (Norris and Evans 2000; Sheldon and Verhulst 1996). Considering the particular case of immune genes, the risk of immunopathology has been proposed as a universal energy-independent cost of resistance.

12.4.2.2 Risks of Immunopathology

Immune response activation may induce great damages to the hosts themselves (Graham et al. 2005; Pennington et al. 2009). It may contribute to the deleterious effects observed in severe parasite-related infectious diseases. For example in mice, 10% of malarial anemia is explained by immunological exuberance rather than parasite-mediated destruction of red blood cells (Graham et al. 2005). It also concerns the development of autoimmune diseases observed in humans and vertebrates (see Sadd and Schmid-Hempel 2009; Sorci and Faivre 2009)

Many demonstrations of the impact of such risk on host resistance gene polymorphism concern *Mhc* genes. Indeed, examination of *Mhc* genetic diversity has revealed that the considerable polymorphism observed is, at least partly, shaped by a trade-off between selection pressures exerted by parasites and auto-immune disease risks (Maizels 2009). It results that the different alleles that hosts carry could be quantitatively and qualitatively limited by the risk of self-destructive reactions (Wucherpfennig and Strominger 1995). Such optimal diversity at *Mhc* genes has been observed in sticklebacks (Wegner et al. 2003, 2004). Other genetic mechanisms underlying these risks of immunopathology have been identified. For example, *Nramp1*, a gene involved in macrophage activation, increases resistance to tuberculosis but is associated with autoimmune diseases (Searle and Blackwell 1999).

Another investigated aspect of immunopatholoy risks influencing resistance polymorphism concerns the immune inflammatory response and its regulation (Sorci and Faivre 2009). Geographical variability in the expression of genes encoding this response has been demonstrated in humans (Pennington et al. 2009) and rodents (Guivier et al. 2010). Besides, the negative impact of the over-expression of such genes on host fitness has also been well demonstrated (Sell 2001). Such

polymorphism therefore seems to result from the interplay between parasite risk exposure and the risks of immune disorders. But further investigations considering the role of parasite immune evasion genes and the evolution of immune regulation could provide a better understanding of the polymorphism observed at inflammatory response genes.

Different reasons have been proposed to explain why immunopathological risks have not been eliminated during evolution (Graham et al. 2005). Briefly, (i) the cost/benefit balance of fighting parasites could maintain immunopathology: high immune response levels might be selected to ensure parasite control despite the risk of immunopathology, (ii) immunopathology could be a consequence of antagonistic immune pathway regulation (e.g. Th1 vs Th2), (iii) selection may not have had the time to act if immunopathology results from novel parasite infections or infections occurring late in life, and (iv) a positive association between immunopathology and parasite transmission might favour parasite genotypes inducing immunopathology. These hypotheses remain to be investigated. As immunopathology is one of the clearest cost of resistance to parasite infections, such studies will bring important information about the evolution of resistance and of immunogenetic polymorphism in natural populations (Graham et al. 2005).

12.4.3 The Importance of Spatialisation and Stochasticity

Because the fitness costs associated with resistance seemed insignificant in some cases (Coustau et al. 2000), population genetists have questioned whether there was any need for costs of resistance to explain the maintenance of polymorphism in GFG systems. These last 10 years, theoretical and empirical researches have shown that spatial structuring of host populations, gradients in environmental conditions or genetic drift could provide relevant mechanisms to explain how resistance polymorphism could be maintained without any cost.

The metapopulation concept (Olivieri et al. 1990) and the geographic mosaic theory of coevolution (Thompson 1994) have provided relevant frameworks to investigate these questions. Briefly, they assume that spatio-temporal variations in the frequency and intensity of selection acting on resistance genes exist among geographically discrete populations, which are more or less connected by migration and experience recurrent extinction/recolonisation events. The strength of selection will be high in some populations (“hot spots”), potentially favoring different trajectories in different populations, and weak or null in others (“cold spots”). This heterogeneous selection might be the result of geographic variation in biotic (infection risk, life history traits) or abiotic (availability of resources, coinfection) features. It might also result from the variable impact of stochastic events (genetic drift) and migration between populations. By consequence, resistance is expected to exhibit significant geographic structure, with genes evolving at different rates and along varied evolutionary trajectories in different populations.

Globally, theoretical studies have emphasized that polymorphism at resistance genes was easier to maintain in spatially structured populations. Metapopulation dynamics

in homogeneous environment does not seem to induce direct frequency-dependence in GFG models and therefore does not lead to the maintenance of stable polymorphism at resistance genes (Laine and Tellier 2008). However, it allows to maintain transient polymorphism for a longer time than in a single population, under the conditions that selection acting on resistance is weak and genetic drift within populations is high (Salathe et al. 2005; Thrall et al. 2002). Furthermore, considering that resistance is encoded by a large number of loci, Thrall and Burdon (2002) or Sasaki (2000) found that the neutral metapopulation processes of dispersal, extinction, recolonisation and drift, could impede the action of selection, what could contribute to maintain transient resistance polymorphisms.

By contrast, metapopulation dynamics in heterogeneous environments has a strong impact on polymorphism (Lenormand 2002). In this context, the scale at which host and parasite dispersal occurs seems to be a critical factor for the maintenance of polymorphism at resistance genes. For example, Thrall and Burdon (2002) have shown that the highest levels of polymorphism were obtained when host and parasite dispersal occurred at a local scale. Interestingly, heterogeneous selection within metapopulation may be mediated through phenotypic plasticity. Some environmental conditions (temperature, availability of resources, coinfection) may affect host genetic expression of resistance genes, therefore leading to spatial variations of parasite-mediated outcomes among populations, what will help to maintain polymorphism in GFG interactions (Laine and Tellier 2008). It is interesting to note that some empirical studies have shown that the influence of phenotypic plasticity might differ between resistance genes (e.g. Bocher et al. 2007). Such phenomenon could potentially underly the different patterns of polymorphism observed at resistance genes (e.g. Barreiro et al. 2005).

Finally, we would like to stress that metapopulation dynamics will further influence the existence of adaptation and maladaptation of hosts and parasites. A large amount of theoretical and empirical researches have explored the conditions of migration and drift that promote host or parasite local adaptation (e.g. Gandon et al. 1996; Gandon and Michalakis 2002; Nuismer et al. 2000). We will not detail these aspects here as they are developed in other parts of this book and in previous reviews (e.g. Deter et al. 2010). We only would like to emphasize that the links between local adaptation and the maintenance of resistance polymorphism remain to be explored (Laine and Tellier 2008).

12.5 Epidemiological Consequences of Resistance Gene Polymorphism

Ultimately, molecular epidemiology aims to improve our understanding of the risks and dynamics of infectious diseases. Surprisingly, despite the obvious importance of this research axis for health policies, the consequences of resistance gene polymorphism for epidemiology remain poorly investigated. This is probably due to the complexity of host-parasite coevolutionary processes, as the relationships between

evolution of anti-parasitic drug resistance and spatial epidemiology, which are 'less complex', have just began to be the focus of theoretical modelling (e.g. Debarre et al. 2007, 2009). We propose that resistance gene polymorphism are closely linked to epidemiology over both micro- and macro-evolutionary scales as (i) it constrains the composition and diversity of the potential host species communities of a given parasite and (ii) it affects critical parameters of infection dynamics, including the transmission rate and the basic reproductive number R_0. These points are discussed below.

12.5.1 Among Species Resistance Polymorphism and Host Species Jumps

Characterising resistance gene polymorphism might help to predict parasite emergence. For examples, viruses that are likely to switch from one host species to another, therefore causing emerging infectious diseases with more or less dramatic consequences (refs in Woolhouse et al. 2005), usually enter host cells via highly phylogenetically conserved host cell receptors (e.g. for foot and mouth disease or rabies viruses, Woolhouse et al. 2002). Pulliam (2008) thus proposed that the whole characterisation of such receptors (identification and distribution among potential host communities) could provide a key tool to predict the outcomes of host parasite specific interactions. In addition, investigating resistance gene polymorphism may provide insights into the relative probability of occurrence of two phenomenons that contribute to emergence: challenging resistance of a new host species *versus* escaping resistance selected for in an existing host. Species jump, and consequently emergence, would be more likely if escaping resistance was more difficult than infecting a new species (e.g. Schneider-Schaulies 2000).

Similarly, we hypothesize that resistance genes evolving under balancing selection and exhibiting trans-specific polymorphism patterns could also mediate host species jumps and emergence of infectious diseases. To our knowledge, this hypothesis has yet not been assessed.

12.5.2 Within Species Resistance Polymorphism and Epidemiological Parameters

It is widely accepted that disease resistance polymorphism that characterizes most of the species may profoundly shape patterns of disease prevalence and incidence (Thrall and Burdon 2003). Field surveys may result in the description of correlations between the distribution of such polymorphism and parasites (e.g. Guivier et al. 2010). They can for example stress that changes in host resistance polymorphism have affected epidemiology (Guivier et al. in press). Mechanisms underlying impacts on epidemiology can even be identified. Among the most obvious examples,

we can cite the loss of *Mhc* gene diversity, which leads to an increased susceptibility to infections and in turn to infectious disease emergence (e.g. Sommer 2005; Woolhouse et al. 2002), or the increase of resistance genotypes within plant culture, which guarantees a lower risk of catastrophic epidemics (Springbett et al. 2003). Field surveys of resistance gene polymorphism may also help to identify genetic characteristics of 'super-spreader' and 'super-shedder' individual hosts that will have important roles for the epidemiology of infectious diseases (Lloyd-Smith et al. 2005).

Experimental evolution is a first way to provide insights into the evolutionary forces mediating such relationships, but this approach can only concern particular organisms exhibiting facilities of breeding and short generation times (e.g. Brockhurst et al. 2007). Therefore, researches on disease resistance in animals are currently strongly dependent on the development of genetic epidemiological models. They are essential to understand the evolution of resistance genes within host populations and to assess its consequences for disease dynamics and persistence. They will also allow to propose predictions that could further be tested using field surveys.

To our knowledge, the only genetic epidemiological models that have been developed to understand the evolution of resistance and disease dynamics in natural populations have considered resistance to anti-parasitic drugs (Debarre et al. 2007, 2009). It seems that the only models referring to disease resistance have been developed in the context of livestock management. These last 10 years, the discoveries of various QTL associated with disease resistance have emphasized the possibilities of improving disease resistance to reduce the transmission of infection in livestocks (Nath et al. 2004). These genetic epidemiological models have first allowed to identify key resistance genes for enhancement programs from their strong predicted impact on infection dynamics (Detilleux 2005; Nath et al. 2004; Nieuwhof et al. 2009). Genes whose variations influence the latent period (e.g. Prp gene for scrapie in sheep, Dawson et al. 1998) or the recovery period (time before parasite abundance decreases to zero) were likely to profoundly change disease dynamic patterns in terms of observed prevalence and parasite burden (Nath et al. 2004; Nieuwhof et al. 2009). They should thus be considered as targets for enhancement programs.

Second, genetic epidemiological models have analysed the proportion of resistant individuals that had to be present within livestocks to protect the whole population from epidemics. MacKenzie and Bishop (1999) have for example shown that the required proportion of resistant animals depended on the infectiousness of the disease.

Different improvements must now be added to such models. On one hand, no spatial context was considered. Models could further be improved by taking into account the potential influence of livestock spatialisation or heterogeneous environments on epidemiological outcomes. On another hand, a particular attention should be paid to the potential links existing between the expression of such resistance genes and tolerance to parasites, as selection for stronger tolerance would not limit the parasite transmission to other non-domestic animals (natural reservoirs). Coevolution of resistance and tolerance could thus be important for the persistence and dynamics of disease over the whole ecosystem.

12.6 Patterns and Processes Affecting Host Resistance Gene Polymorphism During Biological Invasion and Interspecific Hybridisation

Range expansions and interspecific hybridisations have occurred repeatedly in the history of most if not all species, and they are presently accruing at an increasing rate owing to rapid global changes (Hewitt 2000; Mack et al. 2000; Petit and Excoffier 2009). Both phenomenons are often linked because hybridisation generally follows a change in the distributional range of one of the two parental species. Over the recent years, hypotheses and empirical data have accumulated to underline the importance of host-parasite interaction in these contexts. On one hand, host-parasite interactions may strongly influence the outcomes of biological invasion (Prenter et al. 2004) and hybrid zones (Reullier et al. 2006). One the other hand, it has been postulated that biological invasions and hybrid zones may be important hotspots for the evolution of virulence in parasites (Strauss 1994). In the meantime, recent advances in the fields of population genetics have suggested that both invasion and hybridisation may favour the occurrence of genetic revolutions within species (Excoffier et al. 2009) whose consequences on adaptative genetics (and in particular within host-parasite interactions) remains to be explored. Theoretical advances in spatial genetics further outline the difficulties of distinguishing the outcomes of neutral demographic processes and selection events. We thereafter put this different knowledge into perspectives, stressing on the fruitfulness of integrating the recent advances of both disciplines (epidemiology and population genetics) with the hope to stimulate new studies in that directions.

12.6.1 Host-Parasite Interactions in the Context of Biological Invasion

Recent attention has been directed to the role of host-parasite interactions in modulating the outcomes of biological invasion (Prenter et al. 2004). Transmission of parasites from invading to native species has been largely documented, and there are many examples of introduced parasites decimating naïve host populations (Daszak et al. 2000; Harvell et al. 1999). Escape from the deleterious effects of natural enemies is another frequent explanation given for the success of introduced species, a process coined as the *release enemy hypothesis* (Keane and Crawley 2002). Because most parasites reduce host fitness, an invader that leaves parasites behind, and encounter few new parasites, can be advantaged over native species. Reversely, parasite transmission from natives to invaders has been more scarcely documented. However, such new host-parasite associations can occur and their investigation may be crucial to our understanding of invasion success (Carroll 2007; Strauss et al. 2006). Finally, parasites have also been implicated in the indirect mediation of interspecies interactions, primarily through apparent competition

(ie. interspecific competition indirectly mediated by a shared parasite). Among other examples, there is strong support for the suggestion that introduced pheasants *Phasianus colchicus* are implicated in the decline of the partridge *Perdix perdix*, as a result of apparent competition mediated by the caecal nematode *Heterakis gallinarum* (Tompkins et al. 2000).

Contrasting with the important corpus of hypotheses and empirical data emphasizing the role of host-parasite interactions on the outcome of the invasion, few have been done on the outcomes of the invasion on the host-parasite interactions. Invasion has drastic consequences on the genetic structure of the host and of their parasites. The most well known consequence of invasion is the decay in genetic diversities due to successive genetic drifts and founder effects during the invasion (Dlugosh and Parker 2008). Depletion in genetic diversity does not only affect neutral genetics but also functional genetics as evidenced for immune genes in humans (Prugnolle et al. 2005). Based on population genetic theory, molecular models of immunity and empirical studies, population genetic depletion predicts increased susceptibility of populations to emergent parasites. This prediction was confirmed in the Italian agile frog (*Rana latastei*) where the loss of genetic diversity accompanying range expansion is coincident with increased mortality risk from emergent parasites (Pearman and Garner 2005). Experimental studies comparing the susceptibility of frog populations upon the exposure to an emerging strain of *Ranavirus* showed that invasive populations were more susceptible to disease than native populations. Likely higher schistosomiasis, *Schistosoma mansoni*, transmission in the Senegal River Basin (by comparison with that observed in natural habitats in Zimbabwe) was attributed to the low genetic diversity in the invasive intermediate host, *Biomphalaria pfeifferi* (Campbell et al. 2010). These findings suggest that increased susceptibility to novel parasites throughout substantial portions of species ranges may result from processes that parallel the loss of genetic diversity because of multiple bottlenecks, founder effects and inbreeding during range expansion.

Another consequence of invasion is the spread and loss of genetic resistance against infection. The combination of neutral (mutation, drift and migration) and selective processes may produce striking geographic patterns during range expansion as clines in allele frequencies and geographic patches called sectors (see details below in part 12.6.3). Spatial patterns in the genetics of host resistance have been under-investigated until now. However, improvement of technology for the survey of genetic variation and increase in geographic samplings recently gave to human geneticists the opportunity of searching for spatial signature of selection and dispersal in large worldwide population samples (Novembre and Di Rienzo 2009). A demonstrative example comes from the comparative evolutionary histories of thalassaemia (a red-blood-cell disorder in humans) and malaria. Thalassaemias are the most common Mendelian diseases of humans and constitute a major global health problem (Weatherall 2004). They comprise a group of disorders that are caused by defective production of α-globins (encoded by the identical *Hba1* and *Hba2* genes) and β-globins (encoded by *Hbb*). Broadly speaking, homozygous thalassemia results in severe disease or is fatal, whereas heterozygotes are healthy apart from mild anaemia and have a tenfold reduced risk of

Fig. 12.5 The global distribution of malaria and red-blood-cell disorders. Colours indicate the occurrence of malaria: *white* = absence; *light grey* = sporadic; *medium grey* = intermediate; *dark grey* = high malaria risk. Hatching shows the distribution of red-blood-cell disorders (See details in Cooke and Hill 2001)

severe malaria. Malaria then early appeared as the evolutionary driving force behind thalassemia and other haemoglobinopathies (Haldane 1954). The persistence and spread of thalassemian alleles is regarded as the classic paradigm of balanced polymorphisms in human populations. Owing to their high inclusive cost, the spatial distribution of thalassemian alleles is expected to closely match with that of malaria. Indeed, the global geographical distributions of both largely overlap, and thalassemian alleles are maintained at ~10% frequency only in malaria-endemic regions (Fig. 12.5). Moreover there is also good clinical evidence for protection against malaria morbidity, and microepidemiological surveys confirmed the close relationship between malaria and thalassemia (Flint et al. 1986).

Further studies yet revealed a fascinating coevolutionary history between humans and malaria resulting from the combined influences of human migration, genetic drift and local selection. The two major etiologic agents originated fairly recently in different regions, *Plasmodium falciparum* in Africa, and *P. vivax* in Southeast Asia. *P. falciparum* emerged first in Africa 50,000–100,000 years ago, and displayed a first wave of migration out of Africa during the Pleistocen migration of human beings (Cornejo and Escalante 2006). Then, both plasmodium species showed another worldwide expansion about 5,000–10,000 ago, coincident with the spread of agriculture (Armelagos and Harper 2005). Another striking finding was the

unexpectedly high number of allelic variants, which have all increased in frequency within the last 5,000 years, and are patchily distributed. Alleles frequent in West Africa are rare in Southeast Asia and *vice versa*. India provides a good example of a patchy pattern of genetic resistance generated by migration. *HbS* (a widespread thalassemian allele) is found in high frequency in several very distant locations but not in-between where malaria and other haemoglobinopathies are present (Flint et al. 1998). Haplotype analyses revealed that they were very unlikely to have arisen independently in the different locations, while anthropological and archaeological data support the view that the Indian carriers of *HbS* are descendants of a single population, possibly centred close to the Indus valley. It is likely that this group dispersed after invasions from the north of India by people whose descendants now form the majority of the subcontinent's population. The present-day distribution of *HbS* in India is a consequence of migrations that took place within the last 5,000 years.

A last striking observation is the lack of haemoglobinopathies in the New World Amerindians. A likely explanation would be a loss of resistance alleles along the route of human colonisation, either following genetic drift or negative selection during the crossing of large areas in high latitudes where malaria is absent. Malaria would have been recently imported into the New World after the European colonisations (although the evidence to demonstrate this is scant). Collectively these findings illustrate how complex spatial patterns of genetic variation of resistance genes may emerge from the combined effects of selective (cost of resistance and environmental variation) and demographic processes (mutation, migration and genetic drift). They also put forward the striking coexistence of high genetic variation at resistance genes as already noticed before. Adaptive phenotypes in humans have revealed multiple mutations that confer advantage to the same disease. The generality of these findings deserve further studies on natural populations of plants and animals. This challenge can now be addressed thanks to the development of high-throughput genetic methodologies.

12.6.2 Host-Parasite Interactions in the Context of Interspecific Hybridisation

Parasites may drive the dynamics of hybrid zones (Reullier et al. 2006; Wiley et al. 2009) and, reversely, hybridisation has the potential to alter interactions with enemies (Fritz et al. 1999). Hybridisation has been postulated to generate significant evolutionary changes in the genetic architecture of the resistance of hosts and of the virulence of parasites, leading to the idea that hybrid zones may be hotspots of evolution for host-parasite interaction (Strauss 1994).

Many studies have examined how hybridisation may influence host-parasite interactions (Fritz et al. 1999) leading to a great variety of outcomes. Hybrids may show higher, lower, intermediate or similar susceptibility than parental populations. The preferences toward parental or hybrid taxon may change over time in a given site (Wolinska et al. 2008). Hybrids may display higher resistance to a specific

parasite, and lower resistance to another one (Fritz et al. 2003). Generally speaking, the fate of hybrid adaptation to parasites is probably influenced by the genetic basis of the resistance in each parent (single vs. multigenic) and the degree of specialisation of the parasites. In monogenic systems and when genetic variation is additive, resistance of hybrids is expected to be intermediate. Yet in the case of multigenic systems, the outcome of hybridisation may be highly unpredictable, either leading to new combinations of resistance, or alternatively to disruption of well-adapted gene complexes and loss of resistance. Fritz et al. (2003) proposed four alternative scenarios for infection levels in hybrids: the additive scenario set an intermediate resistance of hybrids; the dominance occurs when the hybrid resistance is similar to one of the two parents; the hybrid resistance when hybrids are more resistant that either parents; and the hybrid susceptibility when the hybrids are less resistant that either parents. Alternatively, coevolutionary oscillations may also provide likely explanation for the high variability of hybrid susceptibility to infection (Wolinska et al. 2008). This scenario, which hypothesizes that hybrid resistance varies over time through frequency-dependent selection, was evidenced in natural *Daphnia* populations. Distinguishing between coevolutionary oscillations and the scenario proposed by Fritz and colleagues requires multiple samplings over time, because the pattern observed at a given time of a coevolutionary cycle may easily be referred to one or another Fritz's scenario (Fig. 12.6).

While discussions of host-parasite interactions in hybrid zones have largely focused on genetic consequences, the results presented in most empirical studies have in fact centred on phenotypic patterns of infection, damage or disease occurrence in natural populations. Studies documenting adaptive genetic patterns generated by hybridisation are scarce. Yet hybridisation may favour the transfer of adaptations, as resistance genes, across species. Tompkins et al. (2006) tested the hypothesis that a degree of interspecific hybridisation may improve the viability of endangered species by enriching depauperate gene pools. Controlled hybridisation of an endangered parakeet *Cyanoramphus forbesi* with a widespread relative *C. novaezelandiae* in New Zealand allowed improving the immune function in hybrids (by comparison with *C. forbesi* individuals whose immune function was depleted as currently observed in small, inbred populations). At least one of the immune indicators measured correlates directly with both survivorship and the viability of small populations. Whether such cases occurs in natural populations have not yet been documented but does not seem unrealistic as many immune genes display trans-species polymorphisms, suggesting that introgression of alleles from one species to another one should be facilitated.

12.6.3 Genetic Consequences of Biological Invasion: Theoretical Consideration on Selective and Neutral Processes

Recent spatially explicit simulation studies have led to unexpected and fascinating results about genetic patterns emerging after a range expansion (Excoffier et al. 2009). Outcomes of spatial simulations show that neutral evolutionary forces

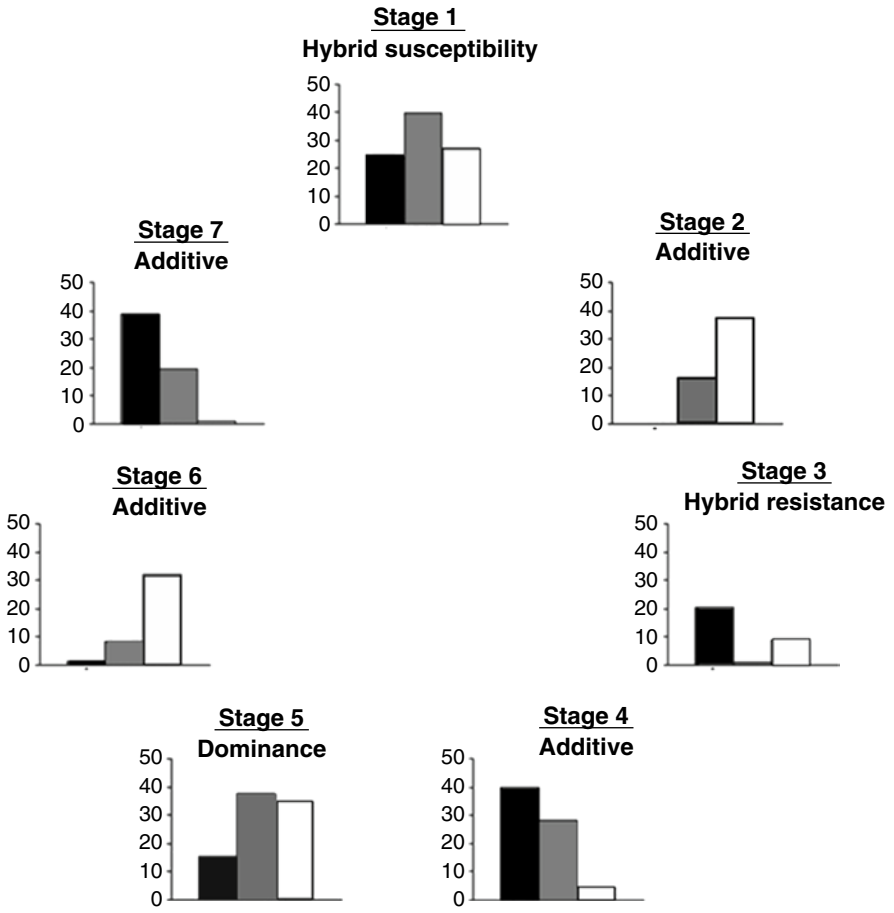

Fig. 12.6 The coevolutionary cycle of host-parasite interactions in hybridising communities. Infection levels are plotted on the *y*-axis. Colours refer to hybrids (*grey*) and parents (*black* and *white*). Hybrids that are over-infected (stage 1) decrease in frequency, and because parasites adapt to the most common taxa, hybrids also decline in relative infection (stages 2 and 3). When hybrids become under-infected, they can ultimately increase in abundance (stages 5–1). At any stage of the coevolutionary cycle, the pattern of infection may be referred to one of the four scenario described by Fritz and colleagues (Details in Wolinska et al. 2008)

(genetic drift, migration and mutation) can alone generate many different patterns that were previously attributed to distinct selective processes. These patterns include allele frequency gradients, surfing of rare variants, high structuring into distinct sectors of low genetic diversity and even massive introgression of local genes into the genome of the invading species. These studies strengthen the potential and pre-eminent role of neutral processes (alone) in shaping adaptive genetics during range expansion. While not yet developed in the context of host-parasite interactions,

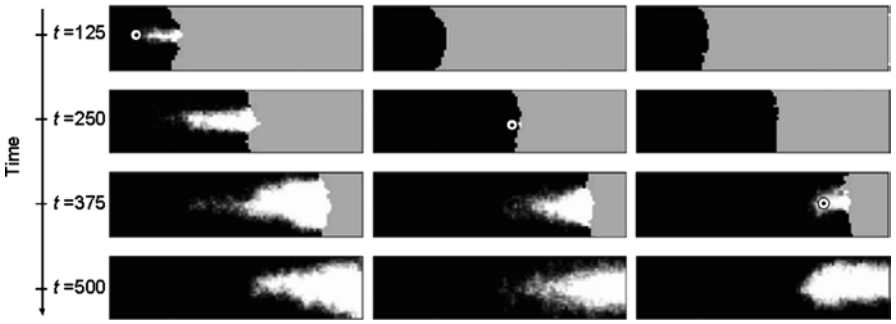

Fig. 12.7 Genetic surfing of alleles in the front of a wave of range expansion. The colonisation starts from a single deme on the left border and advances to the right. At an arbitrary time *t*, a new allele (Figured by a *black dot*) occurs on the edge on the wave and surfs on the expansion wave, reaching high frequencies (*white*). Three cases of successful surfing are reported, that are deleterious, neutral or advantageous alleles. Their dynamics are different: deleterious alleles tend to stay close to the edge and are rapidly lost in the interior, whereas advantageous ones can also travel back and propagate in the already occupied world. Note also that in spite of different modes of selection, and different times of origin, the three kinds of mutations show very similar spatial distribution at time $t = 500$ (Details in Excoffier et al. 2009)

these recent theoretical advances may enlighten our understanding of the dynamics of host resistance genes during biological invasion.

The impacts of selective processes during the invasion dynamics are just at the fringe to be evaluated (Keller and Taylor 2008). Early studies on range expansions showed that advantageous genes should spread at a constant speed into the new environment. However, other kinds of spatial patterns arise because in most cases (1) the environment is not homogeneous (Wegmann et al. 2006); and because the invader may experience (2) long-distance dispersal (Shigesada and Kawasaki 2002) and (3) interspecific competition with resident species (Petit and Excoffier 2009). One of the most impressive features of genetic patterns generated by range expansion is the genetic surfing (Excoffier and Ray 2008). Surfing is the propagation of alleles by the wave front of the invasion (Fig. 12.7). This phenomenon is a result of the intense amount of genetic drift that occurs at the leading edge of a population expansion. Genetic surfing occurs more often in small than in large populations and is favoured by limited dispersal. Travis et al. (2007) demonstrated that deleterious mutation can survive, and even reach high frequencies, close to the wave front while surfing, but should disappear when the colonisation is over. Conditions are then set for maladapation to occur in the wave front. Multilocus maladapted genotypes may persist by surfing during range expansions, be propagated to new environments, and thus promote peak shift in the adaptive landscape (Burton and Travis 2008a). Spatial constraints also highly impact the patterns of genetic variability along the invasion pathway. Some colonisation routes might be more likely than others owing to geographic of ecological barriers. Increasing the environmental heterogeneity may usually reduce the genetic diversity of the invader (Wegmann et al. 2006). Moreover

spatial bottlenecks, which correspond to migration corridors such as land bridges, rivers, coastlines, lead to a much larger probability of surfing during a range expansion, and increase the probability for deleterious mutations to be maintained in the front wave (Burton and Travis 2008b).

Most range expansions do not occur in completely uninhabited areas, and the propagation of a wave thus depends on intraspecifc factors. If interbreeding is possible between a local and an invasive species, a moving hybrid zone develops during the expansion (Barton and Hewitt 1985; Buggs 2007). Simulation and empirical studies show that in most realistic situations, the genome of the invading species should be massively introgressed by that of the local species (Currat et al. 2008). This asymmetrical introgression patterns hold true for a wide range of demographic conditions, including when the fitness of the hybrids is low or moderate (hybrid fitness should only be >10%), or when the competition of the two species drives the local species to extinction. Moreover, a gene introgressing on the wave front can occasionally surf in the invading population and, thus, reach very high frequencies in newly colonised areas. Because the surfing phenomenon is not restricted to neutral variants, deleterious as well as advantageous genes can surf into the invading population.

12.6.4 Prospects on the Fate of Host-Parasite Interactions During Invasion and Hybridisation

Explicit considerations of spatial processes have provided profound and often unexpected insights, revealing that the dynamics of the wave front of invasion may promote genetic revolutions and spatial structures, mimicking adaptative processes (Excoffier and Ray 2008). Although not yet evaluated, consequences on the dynamics of host-parasite interactions are likely and probably not readily predictable. It is now clear that very complex genetic landscape can emerge through a range expansion, leading to a variety of conditions, which may largely influence the outcomes of host-parasite interactions. It is of interest that such outcomes were independently approached by ecologists, which early postulated that biological invasions and hybrid zones could be important hotspots for the evolution of novelties in host-parasite interactions (Strauss 1994). Neutral processes during range expansion may set the conditions for establishing a mosaic of coevolutionary outcomes. For instance, depletion in genetic diversity, but also surfing of maladaptations may lead to local burst in parasites on the road of the invasion. In this context, range expansion could play a major role in modulating the risk for disease emergence, as far as the interaction is mediated by genetic components. This corroborates empirical observations that relate increased risk of disease emergence to global change and biological invasion (e.g. Benmayor et al. 2009; Jones et al. 2008). Determining the part of changes in the genetics of host-parasite interactions that may explains increasing disease emergence deserves further investigations.

Genetic novelty may likely arise from complex interactions between neutral and selective evolutionary forces during the invasion process under non-equilibrium demographic conditions. In some case, a parasite may reach a new area long time after the colonisation by its host. This seems to have been the case for malaria and human invasion of the New World (see above). Another well-documented study examined the rate and magnitude of evolutionary change in an interaction involving coevolved species that were separated and then, by sequential range extensions, reassociated (Zangerl et al. 2008). Within its native area (Europe), the interaction between *Depressaria pastinacella* (parsnip webworm) and wild parsnip (*Pastinaca sativa*) is characterised by chemical phenotype matching, ostensibly mediated by reciprocal selective responses. The reassociation of the webworm, well after the introduction the wild parsnip in New Zealand, revealed a profound alteration of the selection regime, favouring trait remixing and rapid chemical changes in parsnip populations as predicted by the geographic mosaic theory (Zangerl et al. 2008). Finally, another study also documented rapid and radical evolutionary changes in host-parasite interaction for two coevolved species, which were introduced altogether in their non-indigenous range (Gilbert and Parker 2010). Experimental and empirical data collected on clovers and their fungal pathogen showed infection patterns consistent with the hypothesis of adaptive evolution in both the pathogen (ability to infect) and the host (tolerance to infection). Collectively these few studies suggest the potential for rapid evolution to alter interactions between invaders and their natural enemies. To what extent these alterations can be related to genetic changes open new perspectives for future researches.

Finally, where invaders and native hosts can hybridise, there is clearly a potential for transfer of adaptation across species. Hybridisation between locally adapted populations can have significant evolutionary consequences, including the transfer of resistance genes to the invading populations. Surprisingly, theoretical studies suggest that asymmetric gene transfers, from the local to the invasive species, are greatly favoured (Petit and Excoffier 2009). Such rather non-intuitive phenomenon could be an important mechanism allowing alien species to rapidly acquire local adaptations (in particular to natural enemies) to his novel environment.

12.7 Conclusions

Molecular epidemiology of host resistance genes has become the focus of many empirical researches these last years. They have concerned various organisms, evolutionary scales, and mechanisms of resistance. A remarkable aspect of these works is that they have first been investigated independently through different disciplines, including genomics, evolutionary biology of host-parasite interactions or immunology. All these approaches are now closely interacting, therefore providing a multidisciplinary framework that will soon bring new responses to the huge amounts of unsolved questions, which include the patterns of host resistance gene polymorphism

and the processes mediating these patterns. Besides, the strong parallel advances in parasite molecular epidemiology will also allow the development of coevolutionary molecular epidemiology studies. Although it is obvious that they will first concern models of medical or agronomical issues, we have no doubt that fascinating data will soon be available for a broader array of host-parasite systems. This accumulation of molecular epidemiological patterns observed *in natura* is likely to facilitate the development of epidemiological genetic modelling, which is up to now restricted to the context of livestock management.

From this chapter, we would like to emphasize four points that we think deserve particular attention. First, the phenotypic plasticity of resistance that can be mediated through parasite immune modulation. Recent discoveries of parasite driven phenomenon interfering with the expression of host genes, for examples the action of viral microRNAs or parasite molecular mimicry, have opened new perspectives for the study of resistance variability. Evolutionary biologists now need to explore the importance of this parasite-mediated resistance phenotypic plasticity and its consequences for the risks of further co-infections, in particular in the context of potential emerging infectious diseases. Following this idea, we think that the links between the evolution of immune regulation (inherent to the host itself or to immunomodulating parasites), tolerance and immunopathology have hardly been touched upon since the hypotheses proposed by Graham et al. (2005). Hantaviruses could be a relevant model to investigate these questions as they mediate chronic asymptomatic infections in their natural hosts (rodents and insectivores) while inducing more or less severe transient diseases and immunopathologies in humans, that are accidental hosts (see details in Guivier et al. 2010). Whether the patterns observed in natural hosts result from the evolution of tolerance or an immuno-regulation mediated by hantaviruses is an open-question. Thirdly, this chapter has frequently referred to the concepts and results of immunoecology, a recently emerged discipline that aims to explain the variability of immune responses in natural populations (Schmid-Hempel 2003; Sheldon and Verhulst 1996). Advances in the molecular epidemiology of disease resistance should largely contribute to the development of this discipline, which defaults are the over-simplified assays of immune responses and the strong importance given to energy-dependent costs of resistance, at the expense of other, potentially genomic, mechanisms, that could prevent the fixation of resistance in natural populations. Several interesting results have emerged and deserve further attention as they are yet too exploratory or preliminary to be generalised. For example, theoretical modelling developed by Boots and Bowers (2004) suggest that the coexistence of strains or polymorphism between strains with very different degrees of acquired immunity (in terms of length) is very unlikely, while polymorphisms in avoidance strategies or recovery rates are expected to be frequent. We are looking forward further empirical datasets to see whether patterns observed in the wild corroborate or contradict Boots and Bowers' predictions.

Lastly, we have stressed the huge gaps that remained to be assessed to understand the evolution of host resistance gene/polymorphism and adaptation in the situations of biological invasion/range expansion or hybridisation. We were very surprised to find only few studies documenting the dynamics of host-pathogens interactions in the context of biological invasion and interspecific hybridisation,

and particularly the dynamics of genes under selection by this interaction. Because humans are known to have experienced a serial worldwide expansion, and because there is an increasing interest in understanding the heterogeneity of disease incidence and emergence throughout the world, several well-documented studies yet come from the human-disease literature. Owing to recent improvement in technologies allowing rapid, and easy acquiring of large amount of genomic data, we are confident that similar studies will soon be conducted on natural populations of various plants and animals.

References

Abrams PA (2001) Modelling the adaptive dynamics of traits involved in inter- and intraspecific interactions: an assessment of three methods. Ecol Lett 4:166–175

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? Trends Ecol Evol 21:433–438

Agrawal A, Lively CM (2002) Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol Ecol Res 4:79–90

Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and geographical variation at a positively selected MHC class IIB gene of the lesser kestrel (Falco naumanni). Mol Ecol 17:2652–2665

Amato R, Pinelli M, D'Andrea D, Miele G, Nicodemi M, Raiconi G, Cocozza S (2010) A novel approach to simulate gene-environment interactions in complex diseases. BMC Bioinformatics 11:8

Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol Evol 24:21–30

Armelagos GJ, Harper KN (2005) Genomics at the origins of agriculture, part two. Evol Anthropol 14:109–121

Bangham J, Knott SA, Kim KW, Young RS, Jiggins FM (2008) Genetic variation affecting host-parasite interactions: major-effect quantitative trait loci affect the transmission of sigma virus in Drosophila melanogaster. Mol Ecol 17:3800–3807

Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17–30

Barreiro LB, Patin E, Neyrolles O (2005) The heritage of pathogen pressures and ancient demography in the human innate-immunity CD209/CD209L region. Am J Hum Genet 77:869–886

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. Annu Rev Ecol Syst 16:113–148

Benmayor R, Hodgson DJ, Perron GG, Buckling A (2009) Host mixing and disease emergence. Curr Biol 19:764–767

Bergelson J, Dwyer G, Emerson JJ (2001) Models and data on plant-enemy coevolution. Annu Rev Genet 35:469–499

Berggren KT, Ellegren H, Hewitt G, Seddon JM (2005) Understanding the phylogeographic patterns of European hedgehogs, Erinaceus concolor and E. europaeus using the MHC. Heredity 95:84–90

Berthier K, Charbonnel N, Galan M, Chaval Y, Cosson J-F (2006) Migration and recovery of the genetic diversity during the increasing density phase in cyclic vole populations. Mol Ecol 15:2665–2676

Best A, White A, Boots M (2008) Maintenance of host variation in tolerance to pathogens and parasites. Proc Natl Acad Sci USA 105:20786–20791

Bishop SC, Chesnais J, Stear MJ (2002) Breeding for disease resistance: issues and opportunities. In: Proceedings of the 7th world congress on genetics applied to livestock production, Montpellier, France, August 2002. Session 13, pp 1–8

Bocher A, Tirard C, Doums C (2007) Phenotypic plasticity of immune defence linked with foraging activity in the ant Cataglyphis velox. J Evol Biol 20:2228–2234
Boots M, Bowers RG (2004) The evolution of resistance through costly acquired immunity. Proc Biol Sci Lond B 271:715–723
Boots M, Best A, Miller MR, White A (2009) The role of ecological feedbacks in the evolution of host defence: what does theory tell us? Phil Trans R Soc Lond B 364:27–36
Boyce WM, Hedrick PW, Muggli-Cockett NE, Kalinowski S, Penedo MC, Ramey RR 2nd (1997) Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. Genetics 145:421–433
Bradbury J (2004) Ancient footsteps in our genes: evolution and human disease. Gene variants selected during evolution may underlie many common diseases. Lancet 363:952–953
Brockhurst MA, Buckling A, Poullain V, Hochberg ME (2007) The impact of migration from parasite-free patches on antagonistic host-parasite coevolution. Evolution 61:1238–1243
Brown JK (2003) A cost of disease resistance: paradigm or peculiarity? Trends Genet 19:667–671
Bryja J, Galan M, Charbonnel N, Cosson JF (2005) Analysis of MHC class II gene in voles using CE-SSCP. Mol Ecol Notes 5:273–276
Bryja J, Charbonnel N, Berthier K, Galan M, Cosson JF (2007) Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. Mol Ecol 16:5084–5097
Buggs RJ (2007) Empirical study of hybrid zone movement. Heredity 99:301–312
Burton OJ, Travis JM (2008a) The frequency of fitness peak shifts is increased at expanding range margins due to mutation surfing. Genetics 179:941–950
Burton OJ, Travis JM (2008b) Landscape structure and boundary effects determine the fate of mutations occurring during range expansions. Heredity 101:329–340
Campbell G, Noble LR, Rollinson D, Southgate VR, Webster JP, Jones CS (2010) Low genetic diversity in a snail intermediate host (Biomphalaria pfeifferi Krass, 1848) and schistosomiasis transmission in the Senegal River Basin. Mol Ecol 19:241–256
Carius HJ, Little TJ, Ebert D (2001) Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. Evolution 55:1136–1145
Carroll SP (2007) Brave new world: the epistatic foundations of natives adapting to invaders. Genetica 129:193–204
Charbonnel N, Gouy de Bellocq J, Morand S (2006) Immunogenetics of micromammal-macroparasite interactions. In: Morand S, Krasnov B, Poulin R (eds) Micromammals and macroparasites: from evolutionary ecology to management. Springer, Tokyo, pp 401–442
Charbonnel N, Bryja J, Galan M, Deter J, Tollenaere C, Chaval Y, Morand S, Cosson JF (2010) Negative relationships between cellular immune response, Mhc class II heterozygosity and secondary sexual trait in the montane water vole. Evol Appl 3(3):279–290
Coltman DW, Pilkington JG, Smith JA, Pemberton JM (1999) Parasite-mediated selection against inbred Soay sheep in a free-living, island population. Evolution 53:1259–1267
Cooke GS, Hill AVS (2001) Genetics of susceptibility to human infectious disease. Nat Rev Genet 2:967–977
Cornejo OE, Escalante AA (2006) The origin and age of *Plasmodium vivax*. Trends Parasitol 22:558–563
Coscoy L (2007) Immune evasion by Kaposi's sarcoma-associated herpesvirus. Nat Rev Immunol 7:391–401
Cotter SC, Kruuk LE, Wilson K (2004) Costs of resistance: genetic correlations and potential trade-offs in an insect immune system. J Evol Biol 17:421–429
Coustau C, Chevillon C, ffrench-Constant R (2000) Resistance to xenobiotics and parasites: can we count the cost? Trends Ecol Evol 15:378–383
Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920
Daszak P, Cunningham AA, Hyatt AD (2000) Emerging infectious diseases of wildlife – threats to biodiversity and human health. Science 287:443–449

Dawson M, Hoinville LJ, Hosie BD, Hunter N (1998) Guidance on the use of PrP genotyping as an aid to the control of clinical scrapie. Scrapie Information Group. Vet Rec 142:623–625
Debarre F, Bonhoeffer S, Regoes RR (2007) The effect of population structure on the emergence of drug resistance during influenza pandemics. J R Soc Interface 4:893–906
Debarre F, Lenormand T, Gandon S (2009) Evolutionary epidemiology of drug-resistance in space. PLoS Comput Biol 5:e1000337
Deter J, Charbonnel N, Cosson JF (2010) Evolutionary landscape epidemiology. In: Morand S, Krasnov B (eds) The biogeography of interactions. Oxford University Press, New York
Detilleux JC (2005) Genetic management of infectious diseases: a heterogeneous epidemio-genetic model illustrated with S. aureus mastitis. Genet Sel Evol 37:437–453
Devevey G, Christe P (2009) Flea infestation reduces the life span of the common vole. Parasitology 136:1351–1355
Dlugosh KM, Parker IM (2008) Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431–449
Duffy MA, Sivars-Becker L (2007) Rapid evolution and ecological host-parasite dynamics. Ecol Lett 10:44–53
Duncan AB, Little TJ (2007) Parasite-driven genetic change in a natural population of Daphnia. Evolution 61:796–803
Edwards SV, Chesnut K, Satta Y, Wakeland EK (1997) Ancestral polymorphism of Mhc class II genes in mice: implications for balancing selection and the mammalian molecular clock. Genetics 146:655–668
Eisen DP, Minchinton RM (2003) Impact of mannose-binding lectin on susceptibility to infectious diseases. Clin Infect Dis 37:1496–1505
Excoffier L, Ray N (2008) Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol Evol 23:347–351
Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annu Rev Ecol Evol Syst 40:481–501
Ferguson W, Dvora S, Gallo J, Orth A, Boissinot S (2008) Long-term balancing selection at the West Nile virus resistance gene, Oas1b, maintains transspecific polymorphisms in the house mouse. Mol Biol Evol 25:1609–1618
Figueroa F, Gunther E, Klein J (1988) Mhc polymorphism predating speciation. Nature 335:265–267
Flint J, Hill AVS, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MP, Boyce AJ, Weatherall DJ, Clegg JB (1986) High frequencies of α-thalassaemia are the result of natural selection by malaria. Nature 321:744–750
Flint J, Harding RM, Boycec AJ, Clegg JB (1998) The population genetics of the haemoglobinopathies. Clinic Haemat 11:1–51
Flor HH (1956) The complementary genetic systems in flax and flax rust. Adv Genet 8:29–54
Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence handicap. Am Nat 139:603–622
Frank SA (1994) Recognition polymorphism in host–parasite genetics. Phil Trans R Soc Lond B 346:283–293
Fritz RS, Moulia C, Newcombe G (1999) Resistance of hybrid plants and animals to herbivores, pathogens, and parasites. Annu Rev Ecol Syst 30:565–591
Fritz RS, Crabb BA, Hochwender CG (2003) Preference and performance of a gall-inducing sawfly: plant vigor, sex, gall traits and phenology. Oikos 102:601–613
Galvani AP (2003) Epidemiology meets evolutionary ecology. Trends Ecol Evol 18:132–139
Gandon S, Michalakis Y (2000) Evolution of parasite virulence against qualitative or quantitative host resistance. Proc R Soc Lond B 267:985–990
Gandon S, Michalakis Y (2002) Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. J Evol Biol 15:451–462
Gandon S, Capowiez Y, Dubois Y, Michalakis Y, Olivieri I (1996) Local adaptation and gene for gene evolution in a metapopulation model. Proc R Soc Lond B 263:1003–1009

Garland T, Kelly SA (2006) Phenotypic plasticity and experimental evolution. J Exp Med 209:2344–2361
Gautier M, Flori L, Riebler A, Jaffrezic F, Laloe D, Gut I, Moazami-Goudarzi K, Foulley JL (2009) A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. BMC Genomics 10:550 (21 November 2009)
Gavrilovskaya IN, Shepley M, Shaw R, Ginsberg MH, Mackow ER (1998) Beta3 Integrins mediate the cellular entry of hantaviruses that cause respiratory failure. Proc Natl Acad Sci USA 95:7074–7079
Gavrilovskaya IN, Brown EJ, Ginsberg MH, Mackow ER (1999) Cellular entry of hantaviruses which cause hemorrhagic fever with renal syndrome is mediated by beta3 integrins. J Virol 73:3951–3959
Geimonen E, Neff S, Raymond T, Kocer SS, Gavrilovskaya IN, Mackow ER (2002) Pathogenic and nonpathogenic hantaviruses differentially regulate endothelial cell responses. Proc Natl Acad Sci USA 99:13837–13842
Gilbert GS, Parker IM (2010) Rapid evolution in a plant-pathogen interaction and the consequences for introduced host species. Evol Appl 3:144–156
Giraudoux P, Delattre P, Habert M, Quéré J-P, Deblay S, Defaut R, Duhamel R, Moissenet MF, Salvi D, Truchetet D (1997) Population dynamics of fossorial water vole (*Arvicola terrestris scherman*): a land use and landscape perspective. Agric Ecosyst Environ 66:47–60
Gouy de Bellocq J, Charbonnel N, Morand S (2008) Coevolutionary relationship between helminth diversity and MHC class II polymorphism in rodents ? J Evol Biol 21:1144–1150
Graham AL, Allen JE, Read AF (2005) Evolutionary causes and consequences of immunopathology. Annu Rev Ecol Evol Syst 36:373–397
Grenfell BT, Pybus OG, Gog JR, Wood JL, Daly JM, Mumford JA, Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332
Guivier E, Galan M, Chaval Y, Xuereb A, Ribas Salvador A, Charbonnel N, Cosson JF (in press) Landscape genetics highlight the role of metapopulation dynamics of the rodent host in the epidemiology of Puumala hantavirus. Mol Ecol
Guivier E, Galan M, Ribas Salvador A, Xuéreb A, Chaval Y, Olsson G, Essbauer S, Henttonen H, Voutilainen L, Cosson JF, Charbonnel N (2010) *Tnf*-α expression and promoter sequences reflect the balance of tolerance/resistance to *Puumala virus* infection in European bank vole populations. Inf Gen Evol 10:1208–1217
Guivier E, Ribas Salvador A, Chaval Y, Xuéreb A, Henttonen H, Voutilainen L, Poulle ML, Cadet P, Cosson JF, Charbonnel N (2011) Concomitant influence of helminth infection and landscape on the distribution of Puumala hantavirus in its reservoir, Myodes glareolus. BMC Microbiol 11:30
Haldane JBS (1954) An exact test for randomness of mating. J Genet 52:631–635
Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaptation to resist parasites (a review). Proc Natl Acad Sci USA 87:3566–3573
Harvell CD, Kim K, Burkholder JM, Colwell RR, Epstein PR, Grimes DJ, Hofmann EE, Lipp EK, Osterhaus AD, Overstreet RM, Porter JW, Smith GW, Vasta GR (1999) Emerging marine diseases–climate links and anthropogenic factors. Science 285:1505–1510
Hedrick PW (1974) Genetic variation in a heterogeneous environment. I. Temporal heterogeneity and the absolute dominance model. Genetics 78:757–770
Hedrick PW (1976) Genetic variation in a heterogeneous environment. II. Temporal heterogeneity and directional selection. Genetics 84:145–157
Hewitt GM (2000) The genetic legacy of the quaternary ice ages. Nature 405:907–913
Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. Nat Rev Genet 2:516–527
Hunter DJ (2005) Gene-environment interactions in human diseases. Nat Rev Genet 6:287–298
Jarosz B (1991) Host–pathogen interactions in natural populations of *Linum marginale* and *Melampsora lini*: II. Local and regional variation in patterns of resistance and spatial structure. Evolution 45:1618–1627

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P (2008) Global trends in emerging infectious diseases. Nature 451:990–993
Keane RM, Crawley MJ (2002) Exotic plant invasions and the enemy release hypothesis. Trends Ecol Evol 17:164–170
Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion: separating stochastic phenotypic evolution from response to selection. Ecol Lett 11:852–866
King DA, Peckham C, Waage JK, Brownlie J, Woolhouse ME (2006) Epidemiology. Infectious diseases: preparing for the future. Science 313:1392–1393
Koutsogiannouli EA, Moutou KA, Sarafidou T, Stamatis C, Spyrou V, Mamuris Z (2009) Major histocompatibility complex variation at class II Dqa locus in the brown hare (Lepus europaeus). Mol Ecol 18:4631–4649
Kraaijeveld AR, Godfray HCJ (1997) Trade-of between parasitoid resistance and larval competitive ability in *Drosophila melanogaster*. Nature 389:278–280
Laine AL, Tellier A (2008) Heterogeneous selection promotes maintenance of polymorphism in host-parasite interactions. Oikos 117:1281–1288
Landry C, Bernatchez L (2001) Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (Salmo salar). Mol Ecol 10:2525–2539
Langefors ÅH (2005) Adaptive and neutral genetic variation and colonization history of Atlantic Salmon, *Salmo salar*. Env Biol Fish 74:297–308
Lenormand T (2002) Gene flow and the limits to natural selection. Trends Ecol Evol 17:183–189
Li Y, Chan EY, Li JN, Ni C, Peng XX, Rosenzweig E, Tumpey TM, Katze MG (2010) MicroRNA expression and virulence in pandemic influenza virus-infected mice. J Virol 84:3023–3032
Lingappa J, Kuffner T, Tappero J, Whitworth W, Mize A, Kaiser R, McNicholl J (2004) HLA-DQ6 and ingestion of contaminated water: possible gene-environment interaction in an outbreak of leptospirosis. Genes Immun 5:197–202
Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM (2005) Superspreading and the effect of individual variation on disease emergence. Nature 438:355–359
Mack RN, Simberloff D, Lonsdale M, Evans H, Clout M, Bazzaz F (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Issues Ecol 10:689–710
MacKenzie K, Bishop SC (1999) A discrete-time epidemiological model to quantify selection for disease resistance. Anim Sci 69:543–551
Maizels RM (2009) Parasite immunomodulation and polymorphisms of the immune system. J Biol 8:62
Male PJ, Martin JF, Galan M, Bryja J, Deffontaine V, Cosson JF, Michaux J, Charbonnel N (In revision) *Mhc* vs *cytochrome b* evolutionary and phylogeographical patterns in the bank vole *Myodes glarolus* over its European distribution. Biol J Linnean Soc
McKean KA, Yourth CP, Lazzaro BP, Clark AG (2008) The evolutionary costs of immunological maintenance and deployment. BMC Evol Biol 8:76
Miller MR, White A, Boots M (2005) The evolution of host resistance: tolerance and control as distinct strategies. J Theor Biol 236:198–207
Morens DM, Folkers GK, Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. Nature 430:242–249
Moret Y, Schmid-Hempel P (2000) Survival for immunity: the price of immune system activation for bumblebee workers. Science 290:1166–1168
Mou DL, Wang YP, Huang CX, Li GY, Pan L, Yang WS, Bai XF (2006) Cellular entry of Hantaan virus A9 strain: specific interactions with beta3 integrins and a novel 70 kDa protein. Biochem Biophys Res Commun 339:611–617
Nath M, Woolliams JA, Bishop SC (2004) Identifying critical parameters in the dynamics and control of microparasite infection using a stochastic epidemiological model. J Anim Sci 82:384–396
Newman RM, Hall L, Connole M et al (2006) Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5alpha. Proc Natl Acad Sci USA 103:19134–19139

Nieuwhof GJ, Conington J, Bishop SC (2009) A genetic epidemiological model to describe resistance to an endemic bacterial disease in livestock: application to footrot in sheep. Genet Sel Evol 41:19
Nordling D, Andersson M, Zohari S, Gustafsson L (1998) Reproductive effort reduces specific immune response and parasite resistance. Proc R Soc Lond B 265:1291–1298
Norris K, Evans R (2000) Ecological immunology: life history trade-offs and immune defense in birds. Behav Ecol 11:19–26
Novembre J, Di Rienzo A (2009) Spatial patterns of variation due to natural selection in humans. Nat Rev Genet 10:745–755
Nuismer SL, Thompson JN, Gomulkiewicz R (2000) Coevolutionary clines across selection mosaics. Evolution 54:1102–1115
Olivieri I, Couvet D, Gouyon PH (1990) The genetics of transient populations: research at the metapopulation level. Trends Ecol Evol 5:207–210
Ostfeld RS, Glass GE, Keesing F (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. Trends Ecol Evol 20:328–336
Pagès M, Tatard C, Galan M, Kakko C, Chaval Y, Cosson JF, Charbonnel N (Submitted) Genetic variation within β3 subunit integrin sequences and rodent species susceptibility to pathogenic hantaviruses. J Gen Virol
Pearman PB, Garner TWJ (2005) Susceptibility of Italian agile frog populations to an emerging strain of Ranavirus parallels population genetic diversity. Ecol Lett 8:401–408
Pennington R, Gatenbee C, Kennedy B, Harpending H, Cochran G (2009) Group differences in proneness to inflammation. Infect Genet Evol 9:1371–1380
Petit RJ, Excoffier L (2009) Gene flow and species delimitation. Trends Ecol Evol 24:386–393
Prenter J, Macneil C, Dick JT, Dunn AM (2004) Roles of parasites in animal invasions. Trends Ecol Evol 19:385–390
Prugnolle F, Manica A, Charpentier M, Guégan J-F, Guernier V, Balloux F (2005) Pathogen-driven selection and worldwide HLA Class I diversity. Curr Biol 15:1022–1027
Pulliam JR (2008) Viral host jumps: moving toward a predictive framework. EcoHealth 5:80–91
Quintana-Murci L, Alexandre Alcaïs A, Abel L, Casanova J-L (2007) Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. Nat Immunol 8:1165–1171
Raberg L, Sim D, Read AF (2007) Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. Science 318:812–814
Raberg L, Graham AL, Read AF (2009) Decomposing health: tolerance and resistance to parasites in animals. Phil Trans R Soc Lond B 364:37–49
Rausher MD (2001) Co-evolution and plant resistance to natural enemies. Nature 411:857–864
Restif O (2009) Evolutionary epidemiology 20 years on: challenges and prospects. Infect Genet Evol 9:108–123
Restif O, Koella JC (2004) Concurrent evolution of resistance and tolerance to pathogens. Am Nat 164:E90–E102
Reullier J, Perez-Tris J, Bensch S, Secondi J (2006) Diversity, distribution and exchange of blood parasites meeting at an avian moving contact zone. Mol Ecol 15:753–763
Roy BA, Kirchner JW (2000) Evolutionary dynamics of pathogen resistance and tolerance. Evolution 54:51–63
Sadd BM, Schmid-Hempel P (2009) Principles of ecological immunology. Evol Appl 2:113–121
Salathe M, Scherer A, Bonhoeffer S (2005) Neutral drift and polymorphism in gene-for-gene systems. Ecol Lett 8:925–932
Sasaki A (2000) Host-parasite coevolution in a multilocus gene-for-gene system. Proc R Soc Lond B 267:2183–2188
Saucy F (1994) Density dependence in time series of the fossorial form of the water vole, Arvicola terrestris. Oikos 71:381–392
Schmid-Hempel P (2001) On the evolutionary ecology of host-parasite interactions: addressing the question with regard to bumblebees and their parasites. Naturwissenschaften 88:147–158
Schmid-Hempel P (2003) Variation in immune defence as a question of evolutionary ecology. Proc R Soc Lond B 270:357–366

Schneider-Schaulies J (2000) Cellular receptors for viruses: links to tropism and pathogenesis. J Gen Virol 81:1413–1429
Searle S, Blackwell JM (1999) Evidence for a functional repeat polymorphism in the promoter of the human NRAMP1 gene that correlates with autoimmune versus infectious disease susceptibility. J Med Genet 36:295–299
Seddon JM, Ellegren H (2004) A temporal analysis shows major histocompatibility complex loci in the Scandinavian wolf population are consistent with neutral evolution. Proc R Soc Lond B 271:2283–2291
Sell S (2001) Immunology, immunopathology and immunity. ASM Press, Washington, DC
Sheldon BC, Verhulst S (1996) Ecological immunology: costly parasite defences and trade-offs in evolutionary ecology. Trends Ecol Evol 11:317–321
Shigesada N, Kawasaki K (2002) Invasion and the range expansion of species: effects of long-distance dispersal. In: Bullock JM, Kenward RE, Hails RS (eds) Dispersal ecology. Blackwell Scientific, Oxford, pp 350–373
Simms EL, Triplett J (1994) Costs and benefits of plant-responses to disease – resistance and tolerance. Evolution 48:1973–1985
Simms E, Triplett J (1997) Costs and benefits of plants to disease: resistance and tolerance. Evolution 48:1973–1985
Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:16
Sorci G, Faivre B (2009) Inflammation and oxidative stress in vertebrate host-parasite systems. Phil Trans R Soc Lond B 364:71–83
Springbett AJ, MacKenzie K, Woolliams JA, Bishop SC (2003) The contribution of genetic diversity to the spread of infectious diseases in livestock populations. Genetics 165:1465–1474
Springer YP (2007) Clinal resistance structure and pathogen local adaptation in a serpentine flax-flax rust interaction. Evolution 61:1812–1822
Spurgin LG, Richardson DS (2010) How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc Lond B 277:979–988
Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature 400:667–671
Stanford MM, McFadden G, Karupiah G, Chaudhri G (2007) Immunopathogenesis of poxvirus infections: forecasting the impending storm. Immunol Cell Biol 85:93–102
Stearns S (1992) The evolution of life histories. Oxford University Press, Oxford
Strauss SY (1994) Levels of herbivory and parasitism in host hybrid zones. Trends Ecol Evol 9:209–214
Strauss SY, Webb CO, Salamin N (2006) Exotic taxa less related to native species are more invasive. Proc Nat Acad Sci USA 103(15):5841–58415
Svensson EI, Raberg L (2010) Resistance and tolerance in animal enemy-victim coevolution. Trends Ecol Evol 25(5):267–274
Tellier A, Brown JK (2007) Polymorphism in multilocus host parasite coevolutionary interactions. Genetics 177:1777–1790
Thompson JN (1994) The coevolutionary process. The University of Chicago Press, Chicago
Thrall PH, Burdon JJ (2002) Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. Plant Pathol 51:169–184
Thrall P, Burdon JJ (2003) Evolution of virulence in a plant host-pathogen metapopulation. Science 299:1735–1737
Thrall PH, Burdon JJ, Bever JD (2002) Local adaptation in the Linum marginale-Melampsora lini host-pathogen interaction. Evolution 56:1340–1351
Tollenaere C, Bryja J, Galan M, Cadet P, Deter J, Chaval Y, Berthier K, Ribas Salvador A, Voutilainen L, Laakkonen J, Henttonen H, Cosson JF, Charbonnel N (2008) Multiple parasites mediate balancing selection at MHC class II genes: insights from multivariate analyses and population genetics in the fossorial water vole. J Evol Biol 21:1307–1320
Tompkins DM, Greenman JV, Robertson PA, Hudson PJ (2000) The role of shared parasites in the exclusion of wildlife hosts: Heterakis gallinarum in the ring-necked pheasant and the grey partridge. J Anim Ecol 69:829–840

Tompkins DM, Mitchell RA, Bryant DM (2006) Hybridization increases measures of innate and cell-mediated immunity in an endangered bird species. J Anim Ecol 75:559–564
Travis JMJ, Münkemüller T, Burton OJ, Best A, Dytham C, Johst K (2007) Deleterious mutations can surf to high densities on the wave front of an expanding population. Mol Biol Evol 24:2334–2343
Trowsdale J, Parham P (2004) Defense strategies and immunity-related genes. Eur J Immunogenet 34:7–17
Weatherall DJ (2004) History of genetic disease: Thalassaemia: the long road from bedside to genome. Nat Rev Genet 5:625–631
Wedekind C, Walker M, Little TJ (2006) The separate and combined effects of MHC genotype, parasite clone, and host gender on the course of malaria in mice. BMC Genet 7:55
Wegmann D, Currat M, Excoffier L (2006) Molecular diversity after a range expansion in heterogeneous environments. Genetics 174:2009–2020
Wegner KM, Kalbe M, Kurtz J, Reusch TB, Milinski M (2003) Parasite selection for immunogenetic optimality. Science 301:1343
Wegner KM, Kalbe M, Schaschl H, Reusch TB (2004) Parasites and individual major histocompatibility complex diversity–an optimal choice? Microbes Infect 6:1110–1116
Wiley C, Qvarnstrom A, Gustafsson L (2009) Effects of hybridization on the immunity of collared Ficedula albicollis and pied flycatchers F-hypoleuca, and their infection by haemosporidians. J Avian Biol 40:352–357
WHO (2005) The world health report (World health organization, Genève, 2005)
Williams TD, Christians JK, Aiken JJ, Evanson M (1999) Enhanced immune function does not depress reproductive output. Proc R Soc Lond B 266:753–757
Wilson K, Thomas MB, Blanford S, Doggett M, Simpson SJ, Moore SL (2002) Coping with crowds: density-dependent disease resistance in desert locusts. Proc Natl Acad Sci USA 99:5471–5475
Wolinska J, Lively CM, Spaak P (2008) Parasites in hybridizing communities: the red queen again? Trends Parasitol 24:121–126
Woolhouse ME, Webster JP, Domingo E, Charlesworth B, Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet 32:569–577
Woolhouse ME, Haydon DT, Antia R (2005) Emerging pathogens: the epidemiology and evolution of species jumps. Trends Ecol Evol 20:238–244
Wucherpfennig KW, Strominger JL (1995) Selective binding of self peptides to disease-associated major histocompatibility complex (MHC) molecules: a mechanism for MHC-linked susceptibility to human autoimmune diseases. J Exp Med 181:1597–1601
Zangerl AR, Stanley MC, Berenbaum MR (2008) Selection for chemical trait remixing in an invasive weed after reassociation with a coevolved specialist. Proc Natl Acad Sci USA 105:4547–4552

Chapter 13
How Does Biodiversity Influence the Ecology of Infectious Disease?

Gwenaël Vourc'h, Olivier Plantard, and Serge Morand

Abstract Over the past years, biodiversity has been reduced on an unprecedented scale, while new infectious diseases are emerging at an increasing rate. Greater overall biodiversity could lead to a greater diversity of hosts and thus of pathogens. Yet disease regulation – due to the buffering role of host diversity – is considered to be one of the services provided by biodiversity. In this chapter, we ask how biodiversity is linked to infectious disease. First, we investigate the influence of the biodiversity of pathogens. We highlight that the number of pathogen species is not well known but that new findings are facilitated by the rapid expansion of molecular techniques. We show that, although there is a trend to find higher pathogen richness toward the equator, identifying a global pattern between the richness of all pathogen species and their latitudinal distribution is challenging. We emphasize that pathogen intraspecific diversity is a crucial factor in disease emergence and allows pathogens to adapt to the selective pressures they face. In addition, the selective pressure acting on hosts due to parasite, and reinforced by parasite diversity within hosts seems to be a major evolutionary and ecological force shaping hosts biodiversity. Second, we

G. Vourc'h (✉)
Unité d'Epidémiologie Animale (UR346) Institut National de la Recherche Agronomique (INRA), 63122 St Genès Champanelle, France
e-mail: gvourch@clermont.inra.fr

O. Plantard
INRA, UMR 1300 BiOEpAR (Bio-agression, Epidémiologie et Analyse de Risque en santé animale), B.P. 40706, Nantes F-44307, France

LUNAM Université, Oniris, Nantes F-44307, France
e-mail: olivier.plantard@oniris-nantes.fr

S. Morand
Institut des Sciences de l'Evolution, CNRS-IRD-UM2, CC65, Université Montpellier 2, F-34095 Montpellier, France

CIRAD AGIRs, F-34398 Montpellier, France
e-mail: serge.morand@univ-montp2.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_13,

investigate how the diversity of hosts influences infectious disease ecology. For multi-host diseases, a change in host species richness or abundance can modify the dynamics of local infectious diseases by either reducing ("dilution effect") or increasing ("amplification effect") the risk of transmission to the targeted host species. The underlying hypothesis is that, the competence of reservoirs varies according to the host species. The dilution effect has been demonstrated mainly through theoretical work and there have been only few case studies. Regarding the genetic diversity of host, an important issue is: to what extent does a reduction of this diversity impact the ability of the host population to response to infectious diseases? Third, we rapidly examine the role of biodiversity in the treatment of infectious diseases. To conclude, we consider that the consequences of the loss of species biodiversity on infectious diseases is still largely unknown, notably due to the lack of knowledge on the dynamics of host-pathogen relationships, especially at the population and at the community level.. We highlight that work on multi-host/ulti-pathogen systems should be fostered and that new approaches, such as metagenomic investigations that does not require *a priori* assumptions, are promising to describe a community of pathogens and their interactions.

13.1 Introduction

Over the past 50 years, human activity has altered habitats and reduced biodiversity on an unprecedented scale that comes close to mass extinction (MEA 2005). At the same time, new infectious diseases seem to be emerging at an increasing rate (Wilcox and Gubler 2005). During this period, there has been a dramatic spread of highly pathogenic diseases such as AIDS and multi-drug resistant bacterial infections, and in more recent years SARS, West Nile in North America, and highly pathogenic influenza viruses (Jones et al. 2008).

Habitat loss, largely a result of the conversion of forests and savannas into agricultural land, cities, and industrial sites, is the major cause of change in biodiversity. Biodiversity represents the diversity of life at all levels of biological organization, from the genes within populations, the species that compose a community, to the communities that compose ecosystems. Intuitively, one might assume that greater overall biodiversity would lead to a greater diversity of pathogens and hosts, and thereby increase the incidence of infectious diseases (Dunn et al. 2010). Yet disease regulation is said to be one of the services provided by biodiversity because a high level of species diversity creates a buffer that reduces the risk of transmission (MEA 2005; Walpole et al. 2009). Scientific evidence supporting both of these views is beginning to emerge, but the core question remains: how is biodiversity linked to infectious disease? This is the question addressed in this chapter.

Pathogens are organisms that have a negative impact on the fitness of their host(s), often, if now always, by producing visible symptoms (e.g. a disease). Such trophic interaction between two organisms, a host and a parasite, is just one of several interactions that take place within communities and ecosystems, others being

those of prey-predator and plant-phytophagous for instance (Begon et al. 2006). To date, more attention has been paid to these other interactions, and their roles in ecosystem functioning (e.g. Steffan and Snyder 2010), than to pathogen-host interactions, and food web studies only recently have begun to take parasites into consideration (Arias-González and Morand 2007; Lafferty et al. 2006). Studies incorporating pathogens are scarce (Hudson et al. 2006), probably due to the difficulties of surveying pathogens (using intrusive or even destructive sampling methods…). Moreover, the systematics and even basic aspects of parasite biology often are unknown. However, although numerous species of pathogens still need to be described (Dobson et al. 2008), there is no doubt that pathogens represent a large part of biodiversity on earth. Given that each free living species is host to numerous pathogens, and that pathogens of pathogens also exist (consider, for example, phages that are virus affecting bacteria), several authors believe that pathogens may be the most diverse living group on earth (Windsor 1998).

The link between biodiversity and the ecology of infectious diseases is not simple. In this chapter, we investigate how biodiversity influences the ecology of infectious diseases at the intraspecific level (genetic variability of pathogens and hosts) and at the level of communities (species composition). Although we mainly provide examples from human and animal diseases, we also use some illustrations from plants. We describe patterns of biodiversity and the consequences of changes in biodiversity on the ecology of infectious diseases. Lastly, we rapidly examine the role of biodiversity in the treatment of infectious diseases.

13.2 How Does the Diversity of Pathogens Influence Infectious Disease Ecology?

13.2.1 How Many Pathogens Are There?

We shall consider infectious diseases caused by bacteria, virus, fungi, protozoa and endo-parasites, and exclude from our analysis ecto-parasites that are considered here as disease vectors. In the light of the discussion above, the pathogen status of a given living organism clearly is not a straightforward question (consider, for example, the case of some Rickettsia species that are considered to be not only blood vertebrate pathogens, but also tick symbionts, Perlman et al. 2006). This status is related to the host, and varies with individual hosts and species as well as in space and in time (different hosts, for example, can have different resistance or susceptibility). When pathogens have complex life-cycles, some stages may have a different biology (such as biotrophic or necrotrophic plant pathogens, Morris et al. 2009). Furthermore, horizontal gene transfer is so extensive in bacteria that many microbiologists question the existence of species in bacteria, preferring to consider bacteria as populations that exchange genes. However, the existence of core genes responsible for the maintenance of species-specific phenotypic

clusters is an argument supporting the identification of bacterial species (Riley and Lizotte-Waniewski 2009).

For these reasons, combined with the limited knowledge available of the systematics of many pathogens (Brooks and Hoberg 2001), it is difficult to accurately estimate the number of pathogen species. Estimations of pathogen species richness vary from 10% to 50% of living beings (de Meeûs et al. 1998; Poulin and Morand 2004). In estuaries, the biomass of macro and micro-parasites has been estimated as exceeding that of top predators (Kuris et al. 2008). Although the existence of pathogens has been known for a long time, lists of species only were compiled recently for human and animals (Ashford and Crewe 1998; Cleaveland et al. 2001; Taylor et al. 2001), with an update on human pathogens completed in 2007 (Woolhouse and Gaunt 2007). Approximately 1,400 human pathogens were reported, 616 livestock pathogens (cattle, sheep, goats, pigs and horses), and 374 domestic carnivore pathogens (dogs and cats). No clear figure was given for wildlife (but see the Global Mammal Parasite Database at http://www.mammalparasites.org/). On average, over two new species of human viruses also are discovered each year (Woolhouse et al. 2008). Pathogens affecting humans have received more attention than those affecting other species. If one assumes that other animal species are affected in a proportional manner, huge numbers of pathogens remain to be discovered. Altogether, 70,000, 5,000 and 4,000 of fungi, viruses and bacteria respectively have been described, which represent only 5%, 1%, and 0–1% of the total estimated number of species of fungal, viral and bacterial species. It is difficult to know the number of plant pathogens, but a significantly proportion of the fungal, viral and bacterial species are likely to be plant pathogens (Ingram 1999).

Until recently, many new pathogen discoveries relied on the investigation of atypical symptoms. Today, new findings are facilitated by molecular techniques that render it possible to detect and characterise unculturable pathogens and to investigate the presence of genes and genomes independently of individuals (metagenomics). Although multi-host pathogens are more numerous than single hosts, interactions between pathogens and hosts can evolve towards the specialisation of pathogens on a given host species (Cleaveland et al. 2001; Huyse et al. 2005; Pedersen et al. 2005). Such a specialisation can lead to speciation, *id est* the birth of a new pathogen species. Co-cladogenesis, a process of parallel diversification in hosts and pathogens, also can give birth to numerous pathogen species (Page 2003). Until the development of molecular tools, these species were very difficult to distinguish (cryptic species). Systematic investigations using molecular tools have made it possible, however, to reveal a high diversity of pathogens. For instance, in a systematic inventory of viruses in various vertebrate hosts conducted over a 20 year period in the Central African Republic, 919 different viruses were isolated, including 39 new ones (Saluzzo et al. 2004). Two species of *Plasmodium*, *P. falciparum*, infecting humans, and *P. reichenowi*, infecting chimpanzees, were long considered to be within the clade that includes humans and the great apes. However, recent studies of apes in their natural habitat have revealed a much higher diversity of species infecting great apes; in addition, it has been found that *P. falciparum* also infect gorillas (Liu et al. 2010; Prugnolle et al. 2010) and are at the origin of human malaria.

Metagenomic studies in ecosystems such as human faeces (Zhang et al. 2006) and marine sediments (Breitbart et al. 2004) also have revealed that the majority of viral sequences found did not match in the databanks. Finally, new investigations have been launched to monitor people, animals and animal die-offs in areas where people have a high exposure to wildlife. Generic, broad screening tools will be used to detect pathogen species (Wolfe et al. 2007). To our knowledge, a similar approach has not yet been implemented for pathogens of animals or plants.

13.2.2 Does the Worldwide Distribution of Pathogen Species Diversity Mirror That of Other Organisms?

In addition to the inventories of pathogen biodiversity, scientists have investigated which part of the world holds the highest diversity of pathogen species. Many studies on plants and animals have shown that species richness decreases the further one moves away from the equator. The reasons most likely are linked to the area, energy, time and habitat heterogeneity, and geometric constraints (Gaston and Blackburn 2000). Comparative studies exploring pathogen species richness in the tropics compared to temperate zones are scarce and have produced discrepant results. Guernier et al (2004) studied the worldwide distribution of 229 human pathogens (bacteria, virus, fungi, protozoa, and helminths) according to environmental, demographic and economical factors. They found that parasite species richness decreased with latitude and had a spatially nested organization; i.e. some species are widely distributed and occur in many communities while others have more restricted distributions and occur only in a subset of locations. Such findings were confirmed by the analysis of Dunn et al. (2010), who showed that human pathogen diversity was strongly related to both mammal and bird species richness. Diseases that occur in temperate zones also tend to occur in the tropics, while some tropical diseases are restricted only to the tropics. In primates, the number of protozoa species, which primarily are vector-borne transmitted, increase as one approaches the equator, however, the same trend was not found for viruses and helminths (Nunn et al. 2005). Lindenfors et al (2007), who examined the parasite richness of 980 parasite species and 146 terrestrial carnivore species, found that helminth parasite species richness increased the further away one moved from the equator. The reason for this finding is unknown and may be related to a bias in sampling because carnivores inhabiting areas of industrialized countries in the Northern Hemisphere may have been sampled more intensely. Poulin (1995) and Bordes et al (2010) did not find any correlation between helminth species richness at intra or interspecific levels and latitude. Some studies have shown higher tick species richness at lower latitudes (Cumming 2000). However, this is not the case for flea species, which have been found to have low richness at lower latitudes (Krasnov et al. 2004). A final example is Ichneumonid parasitoid hymenoptera. Although a higher specific host diversity is found in the tropics, the number of species of this parasitoid group is similar in both tropical and temperate regions. It has been hypothesised that this is due to habitat fragmentation

(leading to a lower density of hosts); lack of seasonality (and thus of a host population dynamics with peaks and high density of hosts), or the higher content in toxic compounds of tropical plants and thus in phytophagous insects (the "nasty hypothesis") (Gauld et al. 1992). A meta-analysis of parasite-associated host mortality (Robar et al. 2010) revealed that host mortality risk declined as one moves away from the equator, indicating that, in addition to parasitic load, the effect of parasites on host mortality might be determined by some abiotic factors. Thus, although there is a trend to find higher pathogen richness as we move toward the equator, it is thus challenging to identify a global pattern between the richness of all pathogen species and their latitudinal distribution. However, it should be noted that of the 87 pathogens that have been discovered since 1980, most have a global distribution (Woolhouse and Gaunt 2007).

13.2.3 Pathogen Intraspecific Diversity Is One Factor Favouring Disease Emergence

Pathogens generally are characterised as having higher mutation rates and generation times than those of their hosts (Hamilton et al. 1990). Genetic variability also results from recombination during sexual reproduction of eukaryotic pathogens, and any other genetic exchange mechanisms such as bacterial conjugation or viral recombination. In addition, many animal and plant pathogens use a vector to increase gene flow among populations and to reach a new individual host. This genetic diversity is a crucial factor in disease emergence (Cleaveland et al. 2001) and allows pathogens to adapt to the main selective pressures they face: hosts' immune systems, the need to be transmitted, and treatments or vaccines used to counter infections.

The capacity of some pathogens to genetically diversify facilitates their ability to evade host immune systems. One of the best examples is the Human Immunodeficiency Virus (HIV), which is able to change its appearance faster than the time its takes for the immune system to reply (Drosopoulos et al. 1998). Another example is *P. falciparum*, which generates high levels of variability in genes involved in antigenic variability and virulence (*var* genes) by producing frequent recombination events between heterologous chromosomes (Freitas-Junior et al. 2000). High genetic variation of pathogens also is involved in the interspecies infection process as it facilitates the infection of a broader range of host species, which is another characteristic of emerging pathogens (Cleaveland et al. 2001; Woolhouse and Gowtage-Sequeria 2005). The evolutionary potential of pathogens allows them to respond quickly to the directional selective pressure provided by the massive use of drugs (Palumbi 2001). In areas where selective pressure is important, such as in hospitals, multi-resistant bacteria are very frequent (Levy and Marshall 2004). For bacteria, resistant genes probably originated from environmental organisms with which they shared their ecological niche (Aminov and Mackie 2007). These genes can be transferred between different species of bacteria and even between species that colonize different hosts (Nikolich et al. 1994).

Although vaccination is a major advance of modern medicine, it thus far has contributed to the eradication of only one infectious disease in humans (small pox, www.who.int/mediacentre/factsheets/smallpox/en) and one in cattle, buffalo and wildebeest (rinderpest, Normile 2008). As many vaccines do not totally block transmission, vaccination modifies the selective pressure on pathogens. Depending on how vaccines act on the pathogen, the epidemiology consequences can differ (Gandon and Day 2007). For instance, theoretical work has shown that vaccines that reduce the growth rate or toxicity of pathogens also reduce selection pressure against virulent pathogens, leading to higher intrinsic virulence (Gandon et al. 2001). In the poultry industry, an increase in virulence of avian tumour viruses has followed the use of vaccines that reduce virus growth rates (Witter 1997).

13.2.4 Pathogen Diversity to Which Hosts Are Exposed Influences Host Susceptibility to Disease

Although plants lack an adaptive immune system, through evolution they have developed various strategies to stop plant pathogen infections. An induced or acquired systemic resistance occurs following host recognition of a pathogen, which triggers the production of a hypersensitive reaction (Jones and Dangl 2006). Through this mechanism, the plant provides itself protection from secondary infection in distal tissues, even if the plant faces a pathogen for which it does not have the resistance gene (Durrant and Dong 2004).

The immune system of vertebrates acquires its efficiency by being exposed to a diverse array of pathogens. The striking increase in hygiene standards that began in the early twentieth century has considerably lowered humans' exposure to pathogens, at least in developed countries. The immune response triggered by a pathogen can provide some cross protection against other pathogens (e.g. Huang et al. 2008). A low exposure to a diversity of pathogens has had immediate consequences in decreasing the risk of disease. But could this low exposure also induce evolutionary change in the ability of a host to respond to infection? Due to a trade off between investment in disease resistance and other traits linked to fitness, low exposure could decrease the frequency of resistance down through the generations, setting the stage for a potentially devastating outbreak (Altizer et al. 2003; Graham et al. 2010). Domestic species that are bred to be protected from pathogens might be more susceptible to infectious diseases (Lyles and Dobson 1993). Furthermore, it has been suggested that on islands, where some pathogens may be absent, hosts may have lower immune response abilities (island syndrome) (Lee and Klasing 2004). However, studies that have tested this hypothesis, both using experimental and theoretical approaches, have had contrasting results (Beadell et al. 2007; Hochberg and Møller 2001; Matson 2006).

Infections by different species of pathogens or by different strains of the same species within the same individual host or vector are common (Abbot et al. 2007; Cox 2001). In fact, parasite diversity in hosts seems to be a major evolutionary and

ecological force for hosts (Bordes and Morand 2009). These concomitant infections can trigger cross-effective immune responses between pathogens that are antigenically similar, having thus an impact on the issue of the infection (Lee et al. 2010). An infection also can enhance susceptibility to subsequent infection (Cattadori et al. 2007). In particular, individuals with already are in poor physical condition may be more susceptible to multiple infections (Beldomenico and Begon 2009; Telfer et al. 2008). Furthermore, concomitant infection may allow the exchange of genetic material between strains of a given pathogen species or even between species through horizontal gene transfer (see Sect. 13.2.2 above), allowing the emergence of new virulent strains. An extreme case is one in which a pathogen drives the extinction of a population or species. Such scenarios are rare but do occur, generally due to a conjunction of pathogens and other causes. For instance, the decline of amphibian populations around the world is thought to be linked to a fungal pathogen *Batrachochytrium dendrobatidis* causing Chytridiomycosis (Crawford et al. 2010). Amphibians could have an increased susceptibility to the fungus due to changes in temperature variability (Rohr and Raffel 2010). Another example is the dramatic decline of the native red squirrel in the UK that has been attributed to a combination of direct competition with the grey squirrel and disease-mediated competition because the grey squirrel is a reservoir host of the squirrelpox virus that causes disease in the red squirrel (Tompkins et al. 2002). The local extinction of a host also may have tremendous consequences on an entire ecosystem (see for example the case of the wildebeest /rinderpest interactions in the Serengeti, Holdo et al. 2009).

13.3 How Does the Diversity of Hosts Influence Infectious Disease Ecology?

13.3.1 Change in Host Species Richness Modifies Infectious Disease Risk

A change in species richness or abundance can modify the dynamics of local infectious diseases by either reducing or increasing the risk of transmission to the targeted species. The first outcome has been named, the "dilution effect", the second, the "amplification effect". The term "dilution effect" has conveyed different meanings since its first use in disease ecology literature (see Box 2 in the paper Keesing et al. 2006). The broad definition of the dilution effect refers to "the phenomenon – the net effect – when increased species diversity reduces disease risk" that is produced by a variety of mechanisms ("amplification effect" refers to the opposite phenomenon) (Keesing et al. 2006). This applies to vector-borne and directly transmitted diseases, although the concept of dilution has been developed most with regards to the tick-borne Lyme disease (Allan et al. 2003; LoGiudice et al. 2003, 2008).

The hypothesis underlying the amplification and dilution effect is that for many diseases, the competence of reservoirs, i.e. the ability to become infected and

retransmit the pathogen, varies according to the host species (Haydon et al. 2002). The composition of the host community thus can influence the transmission dynamic of the disease. Similarly, since vectors have different competence to transmit pathogens, the composition of the vector community likely influences transmission dynamics. Different mechanisms are thought to be involved, but they are difficult to differentiate (Begon 2008; Keesing et al. 2006). One is the modification of the encounter rate (when reduced, this corresponds to the "dilution effect" *sensu stricto*). In the presence of species that are poorly competent, the transmission event that should link an infectious individual to a susceptible individual instead links infectious individuals to non-competent individuals. For vector-borne diseases, the increased diversity of a poorly competent host species on which the vector feeds increases the proportion of vector bites that are wasted. For direct transmission, the addition of non competent hosts can decrease transmission if these hosts remove infectious particles (Begon 2008). A second mechanism at work is that a high diversity of host species regulates the abundance of the competent host population. This regulation can be mediated by interspecific competition for limiting resources or by predation upon competent hosts. This typically is the idea that underlies biological controls, where organisms prey upon reservoir hosts, vectors or intermediate hosts (Straub and Snyder 2006). A third mechanism is based on the link between species richness and host mortality. This is the case when predators modify the mortality rate of a host and lower pathogen transmission by feeding on a heavily diseased individual (Packer et al. 2003). Two other mechanisms are cited by Keesing et al (2006), but they are difficult to demonstrate: (i) the modification of recovery when species added to a community facilitate the recovery of infected individuals by, for instance, providing resources, and (ii) the modification of transmission once the contact has occurred, for instance, when adding a species modifies the pathogen load within the host.

The dilution effect has been demonstrated mainly through theoretical work; there have been few case studies. One of the main examples is Lyme disease in the USA that is caused by pathogenic bacteria transmitted by ticks. These ticks feed readily on many species of vertebrates and these species vary in their degrees of reservoir competence. The white-footed mouse (*Peromyscus leucopus*) is thought to be the most competent host and dominates in fragmented forests. In native forests, which harbour a higher diversity of species than fragmented forests, ticks have a higher probability to dilute their bite by feeding on a less competent host (Allan et al. 2003; LoGiudice et al. 2003, 2008). However, such a dilution effect has not been demonstrated in Europe, probably because of the complexity of the disease ecology which involves numerous reservoir host and bacteria species (Halos et al. 2010). Another example is the West Nile virus, where an increased diversity of non passerine birds, which are less competent reservoir hosts compared to passerines, was associated with decreased West Nile virus infection in mosquitoes and humans (Ezenwa et al. 2006; Swaddle and Calos 2008). To date, there have been few examples of directly transmitted diseases, but studies on hantaviruses have shown that higher diversity of small mammals appears to regulate reservoir host populations through competition or predation. High small-mammal diversity also might inhibit intraspecific

aggressive encounters between reservoir hosts that result in hantavirus transmission (Suzán et al. 2009).

In plants, crop diversity can reduce the total burden of disease in agricultural systems. This results from the combined effects of (i) the limitation of pathogen dispersal thanks to the physical barriers provided by the presence of non-host plants (Burdon and Chilvers 1982), (ii) induced systemic resistance, and (iii) competition among pathogens. The efficiency of crop mixtures is linked to the size of the area on which this method is used: a high level of success has been observed in a field trial with susceptible and resistant varieties of rice conducted on a large scale (3,342 ha) in China (Zhu et al. 2000).

Illustrations of amplification effects are typically the consequences of species introduction that radically modifies encounter rates. The added species can introduce new pathogens that infect native hosts (spillover) (Bruemmer et al. 2010) or amplify the circulation of local pathogens (spillback) (Kelly et al. 2009). The introduction of additional species also can provide sources of vector meals and increase vector numbers or activity (Saul 2003). For instance, the introduction of Siberian chipmunks (*Tamias sibiricus*) in suburban forests could increase the risk of Lyme disease because this host seems to be more competent than native hosts (Vourc'h et al. 2007). The introduction of the mosquito *Aedes albopictus* in many parts of the world has facilitated the transmission of the chikungunya virus (Benedict et al. 2007; Charrel et al. 2007).

Theoretical works based on deterministic modelling have looked at the conditions in disease transmission dynamics that are needed for the amplification or the dilution effect to occur (Begon 2008; Dobson 2004). When there is a relationship between the risk of a disease, the abundance of the reservoir host, and the abundance of an additional host, the addition of a species does not necessarily decrease the risk. In the case of Lyme disease, for example, tick abundance mainly is determined by the abundance of deer, which are in fact a non competent reservoir. An increased abundance of deer may reduce infection prevalence when immature ticks are feeding on the deer. At the same time, however, the overall number of adult ticks increase proportionally with the number of deer (Begon 2008). Further research in this field are relying on the modelling of the global community competence of hosts and vectors (Roche 2008).

Scientists and societies are increasingly interested in the dilution effect (MEA 2005) due to the link between habitat disturbance, generalist host characteristics, and their competence in disease transmission. Disturbance seems to favour generalist hosts (hosts that use different types of habitats) (Devictor et al. 2008; Marvier et al. 2004), and these hosts often have a broad geographical distribution (McKinney and Lockwood 1999; Smart et al. 2006). Crucially, these species also seem to have a higher competence reservoir or vector reservoir than species that are not favoured by disturbance (Mills 2006; Molyneux et al. 2008; Vittor et al. 2006). For example, many murid rodents that are recognized hosts of hemorrhagic fever viruses are opportunistic species that seem to be favoured in disturbed environments. The question is whether there is a causal link between a species' generalist and opportunist character and its disease competence. Why are murid species associated with

hemorrhagic fever more generalist than those which are not? Could it be possible that specialist species also carry hemorrhagic fever viruses, only these viruses have not yet been identified? Or is there something intrinsic in opportunistic species that makes them more likely to evolve and maintain hemorrhagic fever viruses (Mills 2006)?

13.3.2 Does a Loss in a Host's Genetic Diversity Weaken Its Ability to Respond to Infectious Diseases?

Only a very small subset of plant and animal species have been domesticated (Diamond 2002). Many species of that small subset, for example, cattle (in animals) and maize (in plants), have seen their genetic diversity considerably reduced for the purpose of intensive production (The Bovine Hapmap Consortium, Matsuoka et al. 2002). In the wild, small populations of endangered species often have a very reduced genetic diversity (Keller and Waller 2002). This then raises the following question: to what extent does a reduction of the genetic diversity in a host species impact the ability of the host population to response to infectious diseases (May 1995)?

Genetic loci associated with the major histocompatibility complex (MHC) plays a key role in the acquired immune response of vertebrates (Altizer et al. 2003). MHC genes code for molecules that recognize foreign peptides (antigens) and display them on the cell surface. When the MHC-protein is displayed, it can be presented to immune cells, such as T lymphocytes or Natural Killer cells, which subsequently can trigger an adaptive immune response. Because MHC genes are faced with an important diversity of antigens, they must themselves be diverse. The measure of the genetic diversity of MHC based on an analysis of polymorphism sequences of MHC among individuals in populations has been widely used in conservation biology as a proxy to estimate the fitness of populations confronted by pathogens (Alcaide et al. 2010; Bernatchez and Landry 2003; Sommer 2005). However, the level of genetic variation at MHC loci results from different evolutionary forces (selection, gene flow, mutation) varying both in space and time in co-evolutionary systems involving both hosts and pathogens, making conservation genetics of non-model organism a challenging task (Stockwell et al. 2003). We already have many examples where low genetic diversity of species has favoured the diffusion of, and/or susceptibility to, pathogens. For example, the low genetic diversity of the Tasmanian devil could be involved in its susceptibility to facial tumor disease (McCallum 2008). The low genetic diversity found in commercially traded bee queens has been hypothesised as being one of the factors explaining colony collapse disorder (Le Conte and Navajas 2008). The problem is even more critical in intensive crops in which disease resistance has relied on the use of a very small number of genes. This selection strategy has proven to be ineffective as pathogens manage to overcome the resistance. For example, the resistance of *Brassica napus* (canola, oilseed rape and colza) to *Leptosphaeria maculans*

(causing the blackleg disease) due to a major resistance gene was overcome in an area covering approximately 50,000 ha in South Australia in a period of 3 years (Sprague et al. 2006).

13.4 The Importance of Biodiversity as a Source of Treatment for Infectious Diseases

Even with advances in synthetic chemistry, which provides many biologically active molecules, pharmaceuticals derived from nature remain an important part of pharmaceutical practice today (Newman et al. 2008). All organisms have developed compounds to protect themselves against infectious diseases and to interact with individuals of their own species or other species (e.g. Rogerio et al. 2010). These molecules, coming from all organisms (bacteria, fungus, animals, plants) in terrestrial, marine and extreme ecosystems, represent an amazing diversity that has been tested in the field for millions of years by involving millions of individuals. However, only a very small subset of plants and marine organisms has been investigated for novel bioactive compounds. Furthermore, it is estimated that less that 1% of bacterial species and only 5% of fungal species are known. Those which have not yet been identified could be sources of novel molecules (Cragg and Newman 2005).

Observations of natural medicine practices used by indigenous people have led to the discovery of many drugs. The most well known and widely used pharmaceuticals are quinine, used as a model to synthesize anti-malarial drugs (chloroquine and mefloquine), and artemisinin, indentified as a potent anti-malarial drug by Chinese scientists (Newman et al. 2008). Animals also are a source of inspiration for drugs against infectious diseases. For instance, compounds of the sponge *Cryptotethya crypta* inspired the elaboration of antiviral medication such as AZT used in the treatment of HIV/AIDS (Cragg and Newman 2005). Observing great apes medicate themselves through the plants they eat also could help to reveal new active compounds (Krief et al. 2004).

13.5 Conclusion

Pathogens constitute a large part of biodiversity on earth and are present in all ecosystems and at all trophic levels, where they have a large impact on ecosystem functioning and on the population dynamics and evolution of their hosts. The recent acceleration of biodiversity loss due to human activities deeply impacts host-pathogen dynamics. Pathogens and hosts form co-evolving systems exercising major selective pressures on each other. Furthermore, the virulence or pathogenicity of a given species depends on its environment – which includes the hosts – that is highly variable in space and time. In such a context, human beings will never be able to completely control or eradicate every pathogen species; rather, we should accept that we must

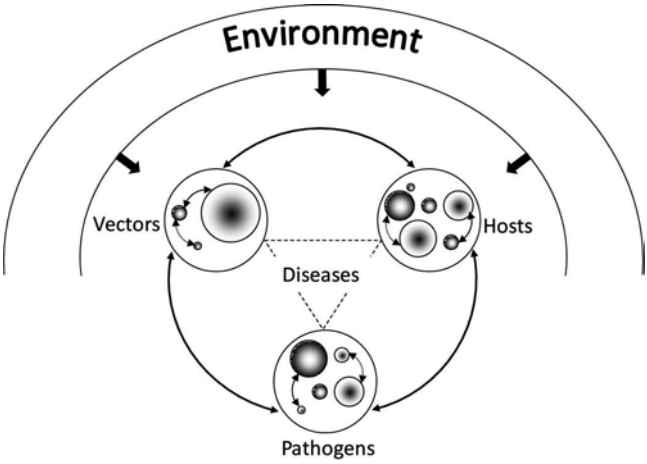

Fig. 13.1 Schematic representation of the link between biodiversity and the ecology of infectious diseases. Diseases results from the complex interactions between the three compartments corresponding to pathogens, hosts and vectors (in the case of vector-borne diseases). The biodiversity of these three compartments can be considered at the community level (each circle corresponding to a species) or at the intraspecific level (each *circle* corresponding to a population or an individual within a population). The *gray shading* off of each unit considered (species, population or individual) illustrates its genetic or phenotypic variability in space and time, while variations in size illustrate frequency or density differences within the ecosystem. Interactions within each compartment can be direct (competition) or indirect (apparent competition…), synergic or antagonistic as illustrated by the different *arrows*

coexist with pathogens. To better understand and predict the evolution of pathogens and the impact of human activities on them, more in-depth studies are needed on how pathogens interact with host communities within different ecosystems.

To understand human, animal, and plant epidemics, two-species systems involving only a single host and a single pathogen species are no longer appropriate. The approach considering multi-host/multi-pathogen systems in their environment is the framework that now needs to be used to deepen our understanding of disease dynamics (Holt et al. 2003; Woolhouse et al. 2001) (Fig. 13.1). However, these dynamics are very complex and difficult to study because precise knowledge regarding the diversity of pathogens and of interactions taking place on several scales is lacking (Lloyd-Smith et al. 2009). In addition to intensive fieldwork to collect adequate data and modeling to understand the main processes, the use of molecular tools in a multi-host/multi-pathogen framework will facilitate investigations into pathogen-host community interactions. In particular, new generation sequencing techniques render it easier to characterize the genetic diversity of pathogens and hosts. For instance, metagenomic investigations allow an approach that does not require *a priori* assumptions that is useful to describe a community of pathogens and their interactions. Molecular techniques also may be used to clarify the taxonomic status of pathogens, revealing cryptic species or host races. With suitable molecular

markers (producing a high level of polymorphism), the analysis of genetic variability within a spatially explicit framework renders it possible to identify the routes followed by a given pathogen. Moreover, molecular techniques can be used to identify genes involved in important life history traits of a pathogen such as virulence and transmission. Better knowledge of the mechanisms involved in host-pathogen interactions, and the extent of their variability, will significantly advance our understanding of outbreaks.

Although our knowledge of the number and variety of pathogens is not complete, it appears that their diversity, like that of their hosts and vectors, is greater in tropical areas than in temperate ones, and in undisturbed habitats than disturbed ones (Chaisiri et al. 2010; Friggens and Beier 2010). The reason we are so concerned by the loss of species biodiversity is because a reduction of biodiversity seems to favour opportunistic species that are highly competent as pathogen reservoirs and vectors. However, this observation was derived from only a few studies and theoretical works, mainly undertaken in the temperate zones. Further investigations should be launched to verify the link between, and understand the process involved in, biodiversity loss and disease dynamics. This especially should be done in the tropics to understand whether high levels of biodiversity create a buffer reducing the risk of disease transmission, and to understand the consequences of biodiversity loss in high biodiversity regions. With global changes, there is a high risk that diseases currently circulating in the tropics will reach temperate zones where species diversity is reduced and the availability of alternate hosts is limited. What could be the consequences of such a shift (Dobson et al. 2006)? What may happen in a world where increased movements of hosts and pathogens, high population densities, and rapidly changing environments increase contact rates, spread, and selective pressures on pathogens and hosts while at the same time a combination of low exposure to pathogen biodiversity and decreased genetic variability in some animals increases susceptibility to new diseases? The investigation of such questions requires collaboration across disciplines and between countries.

References

Abbot P, Aviles AE, Eller L et al (2007) Mixed infections, cryptic diversity, and vector-borne pathogens: evidence from *Polygenis* fleas and *Bartonella* species. Appl Environ Microbiol 73:6045–6052

Alcaide M, Lemus JA, Blanco G et al (2010) MHC diversity and differential exposure to pathogens in kestrels (Aves: Falconidae). Mol Ecol 19:691–705

Allan BF, Keesing F, Ostfeld RS (2003) Effect of forest fragmentation on Lyme disease risk. Conserv Biol 17:267–272

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol Evol 18:589–596

Aminov RI, Mackie RI (2007) Evolution and ecology of antibiotic resistance genes. FEMS Microbiol Lett 271:147–161

Arias-González JE, Morand S (2007) Trophic functioning with parasites: a new insight for ecosystem analysis. Mar Ecol Prog Ser 320:43–53

Ashford RW, Crewe W (1998) The parasites of *Homo sapiens*: an annotated checklist of the protozoa, helminths and arthropods for which we are home. Liverpool School of Tropical Medicine, Liverpool, 728 p

Beadell JS, Atkins C, Cashion E et al (2007) Immunological change in a parasite-impoverished environment: divergent signals from four island taxa. PLoS One 2(9):e896

Begon M (2008) Effect of host diversity on disease dynamics. In: Ostfeld RS, Keesing F, Eviner VT (eds) Effects of ecosystems on disease and of disease on ecosystems. Princeton University Press, Princeton, pp 12–29

Begon M, Townsend CR, Harper JL (2006) Ecology: individuals, populations and communities. Blackwell, Oxford

Beldomenico PM, Begon M (2009) Disease spread, susceptibility and infection intensity: vicious circles? Trends Ecol Evol 25:21–27

Benedict MQ, Levine RS, Hawley WA et al (2007) Spread of the tiger: global risk of invasion by the mosquito *Aedes albopictus*. Vector Borne Zoonotic Dis 7:76–85

Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol 16:363–377

Bordes F, Morand S (2009) Parasite diversity: an overlooked metric of parasite pressures? Oikos 118:801–806

Bordes F, Morand S, Krasnov BR et al (2010) Parasite diversity and latitudinal gradients in terrestrial mammals. In: Morand S, Krasnov BR (eds) The biogeography of host-parasite interactions. Oxford University Press, Oxford, pp 89–98

Breitbart M, Felts B, Kelley S et al (2004) Diversity and population structure of a near-shore marine-sediment viral community. Proc Biol Sci 271:565–574

Brooks DR, Hoberg EP (2001) Parasite systematics in the 21st century: opportunities 303 and obstacles. Trends Parasitol 17:273–275

Bruemmer CM, Rushton SP, Gurnell J et al (2010) Epidemiology of squirrelpox virus in grey squirrels in the UK. Epidemiol Infect 138:941–950

Burdon JJ, Chilvers GA (1982) Host density as a factor in plant-disease ecology. Annu Rev Phytopathol 20:143–166

Cattadori IM, Albert R, Boag B (2007) Variation in host susceptibility and infectiousness generated by co-infection: the myxoma-*Trichostrongylus retortaeformis* case in wild rabbits. J R Soc Interface 4:831–840

Chaisiri K, Chaeychomsri W, Siruntawineti J et al (2010) Human-dominated habitats and helminth parasitism in Southeast Asian murids. Parasitol Res 107:931–937

Charrel RN, de Lamballerie X, Raoult D (2007) Chikungunya outbreaks – the globalization of vectorborne diseases. New Engl J Med 356:769–771

Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Phil Trans R Soc Lond B 356:991–999

Cox FEG (2001) Concomitant infections, parasites and immune responses. Parasitology 122:S23–S38

Cragg GM, Newman DJ (2005) Biodiversity: a continuing source of novel drug leads. Pure Appl Chem 77:7–24

Crawford AJ, Lips KR, Bermingham E (2010) Epidemic disease decimates amphibian abundance, species diversity, and evolutionary history in the highlands of central Panama. Proc Natl Acad Sci USA 107:13777–13782

Cumming GS (2000) Using habitat models to map diversity: pan-African species richness of ticks (Acari: Ixodida). J Biogeogr 27:425–440

de Meeûs T, Michalakis Y, Renaud F (1998) Santa Rosalia revisited: or why are there so many kinds of parasites in 'the garden of earthly delights'? Parasitol Today 14:10–13

Devictor V, Julliard R, Jiguet F (2008) Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. Oikos 117:507–514

Diamond J (2002) Evolution, consequences and future of plant and animal domestication. Nature 418:700–707

Dobson A (2004) Population dynamics of pathogens with multiple host species. Am Nat 164:S64–S78
Dobson A, Cattadori I, Holt RD et al (2006) Sacred cows and sympathetic squirrels: the importance of biological diversity to human health. PLoS Med 3:714–718
Dobson A, Lafferty KD, Kuris AM et al (2008) Homage to Linnaeus: how many parasites? How many hosts? Proc Natl Acad Sci USA 105:11482–11489
Drosopoulos WC, Rezende LF, Wainberg MA et al (1998) Virtues of being faithful: can we limit the genetic variation in human immunodeficiency virus? J Mol Med 76:604–612
Dunn RR, Davies TJ, Harris NC et al (2010) Global drivers of human pathogen richness and prevalence. Proc R Soc Lond B Biol Sci 277:2587–2595
Durrant WE, Dong X (2004) Systemic acquired resistance. Annu Rev Phytopathol 42:185–209
Ezenwa VO, Godsey MS, King RJ et al (2006) Avian diversity and West Nile virus: testing associations between biodiversity and infectious disease risk. Proc R Soc Lond B Biol Sci 273:109–117
Freitas-Junior LH, Bottius E, Pirrit LA et al (2000) Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature 407:1018–1022
Friggens MM, Beier P (2010) Anthropogenic disturbance and the risk of flea-borne disease transmission. Oecologia 164:809–820
Gandon S, Day T (2007) The evolutionary epidemiology of vaccination. J R Soc Interface 4:803–817
Gandon S, Mackinnon MJ, Nee S et al (2001) Imperfect vaccines and the evolution of pathogen virulence. Nature 414:751–756
Gaston KJ, Blackburn TM (2000) Pattern and process in macroecology. Blackwell Science Publications, Oxford, 377 p
Gauld ID, Gaston KJ, Janzen DH (1992) Plant allelochemicals, tritrophic interactions and the anomalous diversity of tropical parasitoids: the 'nasty' host hypothesis. Oikos 65:352–357
Graham AL, Hayward AD, Watt KA et al (2010) Fitness correlates of heritable variation in antibody responsiveness in a wild mammal. Science 330:662–665
Guernier V, Hochberg ME, Guégan J-F (2004) Ecology drives the worldwide distribution of human diseases. PLoS Biol 2:e141
Halos L, Bord S, Cotté V et al (2010) Ecological factors characterizing the prevalence of bacterial tick-borne pathogens in *Ixodes ricinus* ticks in pastures and woodlands. Appl Environ Microbiol 76:4413–4420
Hamilton WD, Axelrod R, Tanese R (1990) Sexual reproduction as an adaptation to resist to parasites (a review). Proc Natl Acad Sci USA 87:3566–3573
Haydon DT, Cleaveland S, Taylor LH et al (2002) Identifying reservoirs of infection: a conceptual and practical challenge. Emerg Infect Dis 8:1468–1473
Hochberg ME, Møller AP (2001) Insularity and adaptation in coupled victim-enemy associations. J Evol Biol 14:539–551
Holdo RM, Sinclair ARE, Dobson AP et al (2009) A disease-mediated trophic cascade in the Serengeti and its implications for ecosystem C. PLoS Biol 7:e1000210
Holt RD, Dobson AP, Begon M et al (2003) Parasite establishment in host communities. Ecol Lett 6:837–842
Huang WJ, Zhang HY, Harrison TJ et al (2008) Cross-protection of hepatitis E virus genotypes 1 and 4 in rhesus macaques. J Med Virol 80:824–832
Hudson PJ, Dobson AP, Lafferty KD (2006) Is a healthy ecosystem one that is rich in parasites? Trends Ecol Evol 21:381–385
Huyse T, Poulin R, Theron A (2005) Speciation in parasites: a population genetic approach. Trends Parasitol 21:469–475
Ingram DS (1999) Biodiversity, plant pathogens and conservation. Plant Pathology 48:433–442
Jones JD, Dangl JL (2006) The plant immune system. Nature 444:323–329
Jones KE, Patel NG, Levy MA et al (2008) Global trends in emerging infectious diseases. Nature 451:990–993

Keesing F, Holt RD, Ostfeld RS (2006) Effects of species diversity on disease risk. Ecol Lett 9:485–498
Keller LF, Waller DM (2002) Inbreeding effects in wild populations. Trends Ecol Evol 17:230–241
Kelly DW, Paterson RA, Townsend CR et al (2009) Parasite spillback: a neglected concept in invasion ecology? Ecology 90:2047–2056
Krasnov BR, Shenbrot GI, Khokhlova IS et al (2004) Flea species richness and parameters of host body, host geography and host 'milieu'. J Anim Ecol 73:1121–1128
Krief S, Martin MT, Grellier P et al (2004) Novel antimalarial compounds isolated in a survey of self-medicative behavior of wild chimpanzees in Uganda. Antimicrob Agents Chemother 48:3196–3199
Kuris AM, Hechinger RF, Shaw JC et al (2008) Ecosystem energetic implications of parasite and free-living biomass in three estuaries. Nature 454:515–518
Lafferty KD, Dobson AP, Kuris AM (2006) Parasites dominate food web links. Proc Natl Acad Sci USA 103:11211–11216
Le Conte Y, Navajas M (2008) Climate change: impact on honey bee populations and diseases. Revue Scientifique et Technique-Office International des Epizooties 27:499–510
Lee KA, Klasing KC (2004) A role for immunology in invasion biology. Trends Ecol Evol 19:523–529
Lee BH, Kim WH, Jeong J et al (2010) Prevalence and cross-immunity of *Eimeria* species on Korean chicken farms. J Vet Med Sci 72:985–989
Levy SB, Marshall B (2004) Antibacterial resistance worldwide: causes, challenges and responses. Nat Med 10:S122–S129
Lindenfors P, Nunn CL, Jones KE et al (2007) Parasite species richness in carnivores: effects of host body mass, latitude, geographical range and population density. Glob Ecol Biogeogr 16:496–509
Liu W, Li Y, Learn GH et al (2010) Origin of the human malaria parasite *Plasmodium falciparum* in gorillas. Nature 467:420–425
Lloyd-Smith JO, George D, Pepin KM et al (2009) Epidemic dynamics at the human-animal interface. Science 326:1362–1367
LoGiudice K, Ostfeld RS, Schmidt KA et al (2003) The ecology of infectious disease: effects of host diversity and community composition on Lyme disease risk. Proc Natl Acad Sci USA 100:567–571
LoGiudice K, Duerr STK, Newhouse MJ et al (2008) Impact of host community composition on Lyme disease risk. Ecology 89:2841–2849
Lyles AM, Dobson AP (1993) Infectious disease and intensive management: population dynamics, threatened hosts, and their parasites. J Zoo Wildlife Med 24:315–326
Marvier M, Kareiva P, Neubert MG (2004) Habitat destruction, fragmentation, and disturbance promote invasion by habitat generalists in a multispecies metapopulation. Risk Anal 24:869–878
Matson KD (2006) Are there differences in immune function between continental and insular birds? Proc R Soc Lond B Biol Sci 273:2267–2274
Matsuoka Y, Vigouroux Y, Goodman MM et al (2002) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084
May RM (1995) Population genetics – the cheetah controversy. Nature 374:309–310
McCallum H (2008) Tasmanian devil facial tumour disease: lessons for conservation biology. Trends Ecol Evol 23:631–637
McKinney ML, Lockwood JL (1999) Biotic homogenization: a few winners replacing many losers in the next mass extinction. Trends Ecol Evol 14:450–453
MEA (2005) Millenium ecosystem assessment. In: Ecosystems and human well–being: biodiversity synthesis. Island Press, Washington, DC
Mills JN (2006) Biodiversity loss and emerging infectious disease: an example from the rodent-borne hemorrhagic fevers. Biodiversity 7:9–17
Molyneux DH, Ostfeld RS, Bernstein A et al (2008) Ecosystem disturbance, biodiversity loss, and human infectious disease. In: Chivian E, Bernstein A (eds) Sustaining life: how human health depends on biodiversity. Oxford University Press, Oxford, pp 287–323

Morris CE, Bardin M, Kinkel LL et al (2009) Expanding the paradigms of plant pathogen life history and evolution of parasitic fitness beyond agricultural boundaries. PLoS Pathog 5:e1000693

Newman DJ, Kilama J, Bernstein A et al (2008) Medicines from nature. In: Chivian E, Bernstein A (eds) Sustaining life: how human health depends on biodiversity. Oxford University Press, New York, pp 117–161

Nikolich MP, Hong G, Shoemaker NB et al (1994) Evidence for natural horizontal transfer of tetQ between bacteria that normally colonize humans and bacteria that normally colonize livestock. Appl Environ Microbiol 60:3255–3260

Normile D (2008) Driven to extinction. Science 319:1606–1609

Nunn CL, Altizer SM, Sechrest W et al (2005) Latitudinal gradients of parasite species richness in primates. Divers Distrib 11:249–256

Packer C, Holt RD, Hudson PJ et al (2003) Keeping the herds healthy and alert: implications of predator control for infectious disease. Ecol Lett 6:797–802

Page RDM (2003) Tangled trees: phylogeny, cospeciation, and coevolution. The University of Chicago Press, Chicago

Palumbi SR (2001) Evolution – humans as the world's greatest evolutionary force. Science 293:1786–1790

Pedersen AB, Altizer S, Poss M et al (2005) Patterns of host specificity and transmission among parasites of wild primates. Int J Parasitol 35:647–657

Perlman SJ, Hunter MS, Zchori-Fein E (2006) The emerging diversity of *Rickettsia*. Proc R Soc Lond B Biol Sci 273:2097–2106

Poulin R (1995) Phylogeny, ecology and the richness of parasite communities in vertebrates. Ecol Monogr 65:283–302

Poulin R, Morand S (2004) Parasite biodiversity. Smithsonian Institution Press, Washington, DC

Prugnolle F, Durand P, Neel C et al (2010) African great apes are natural hosts of multiple related malaria species, including *Plasmodium falciparum*. Proc Natl Acad Sci USA 107:1458–1463

Riley MA, Lizotte-Waniewski M (2009) Population genomics and the bacterial species concept. Methods Mol Biol 532:367–377

Robar N, Burness G, Murray DL (2010) Tropics, trophics and taxonomy: the determinants of parasite-associated host mortality. Oikos 119:1273–1280

Roche B (2008) Complexité des écosystèmes, dynamique de la diversité biologique et maladies infectieuses. Une introduction à l'épidémiologie des communautés. University Montpellier II

Rogerio AP, Sa-Nunes A, Faccioli LH (2010) The activity of medicinal plants and secondary metabolites on eosinophilic inflammation. Pharmacol Res 62:298–307

Rohr JR, Raffel TR (2010) Linking global climate and temperature variability to widespread amphibian declines putatively caused by disease. Proc Natl Acad Sci USA 107:8269–8274

Saluzzo J-F, Vidal P, Gonzalez J-P (2004) Les virus émergents. IRD Editions, Paris, 188 p

Saul A (2003) Zooprophylaxis or zoopotentiation: the outcome of introducing animals on vector transmission is highly dependent on the mosquito mortality while searching. Malar J 2:32–50

Smart SM, Thompson K, Marrs RH et al (2006) Biotic homogenization and changes in species diversity across human-modified ecosystems. Proc R Soc Lond B Biol Sci 273:2659–2665

Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front Zool 2:1–18

Sprague SJ, Marcroft SJ, Hayden HL et al (2006) Major gene resistance to blackleg in Brassica napus overcome within three years of commercial production in southeastern Australia. Plant Dis 90:190–198

Steffan SA, Snyder WE (2010) Cascading diversity effects transmitted exclusively by behavioral interactions. Ecology 91:2242–2252

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. Trends Ecol Evol 18:94–101

Straub CA, Snyder WE (2006) Experimental approaches to understanding the relationship between predator biodiversity and biological control. In: Brodeur J, Boivin G (eds) Trophic and guild in biological interactions control. Springer, Dordrecht, pp 221–239

Suzán G, Marcé E, Giermakowski JT et al (2009) Experimental evidence for reduced rodent diversity causing increased Hantavirus prevalence. PLoS One 4(5):e5461
Swaddle JP, Calos SE (2008) Increased avian diversity is associated with lower incidence of human West Nile infection: observation of the dilution effect. PLoS One 3:e2488
Taylor LH, Latham SM, Woolhouse MEJ (2001) Risk factors for human disease emergence. Phil Trans R Soc Lond B 356:983–989
Telfer S, Birtles R, Bennett M et al (2008) Parasite interactions in natural populations: insights from longitudinal data. Parasitology 135:767–781
Tompkins DM, Sainsbury AW, Nettleton P et al (2002) Parapoxvirus causes a deleterious disease in red squirrels associated with UK population declines. Proc R Soc Lond B Biol Sci 269:529–533
Vittor AY, Gilman RH, Tielsch J et al (2006) The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of *Falciparum* malaria in the Peruvian Amazon. Am J Trop Med Hyg 74:3–11
Vourc'h G, Marmet J, Chassagne M et al (2007) *Borrelia burgdorferi* sensu lato in Siberian chipmunks (*Tamias sibiricus*) introduced in suburban forests in France. Vector Borne Zoonotic Dis 7:637–641
Walpole M, Almond REA, Besancon C et al (2009) Tracking progress toward the 2010 biodiversity target and beyond. Science 325:1503–1504
Wilcox BA, Gubler DJ (2005) Disease ecology and the global emergence of zoonotic pathogens. Environ Health Prev Med 10:263–272
Windsor DA (1998) Most of the species on Earth are parasites. Int J Parasitol 28:1939–1941
Witter RL (1997) Avian tumor viruses: persistent and evolving pathogens. Acta Vet Hung 45:251–266
Wolfe ND, Dunavan CP, Diamond J (2007) Origins of major human infectious diseases. Nature 447:279–283
Woolhouse M, Gaunt E (2007) Ecological origins of novel human pathogens. Crit Rev Microbiol 33:231–242
Woolhouse MEJ, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. Emerg Infect Dis 11:1842–1847
Woolhouse MEJ, Taylor LH, Haydon DT (2001) Population biology of multihost pathogens. Science 292:1109–1112
Woolhouse MEJ, Howey R, Gaunt E et al (2008) Temporal trends in the discovery of human viruses. Proc R Soc Lond B Biol Sci 275:2111–2115
Zhang T, Breitbart M, Lee WH et al (2006) RNA viral community in human feces: prevalence of plant pathogenic viruses. PLoS Biol 4:108–118
Zhu YY, Chen HR, Fan JH et al (2000) Genetic diversity and disease control in rice. Nature 406:718–722

Chapter 14
Epidemiological Interaction at the Wildlife/Livestock/Human Interface: Can We Anticipate Emerging Infectious Diseases in Their Hotspots? A Framework for Understanding Emerging Diseases Processes in Their Hot Spots

Alexandre Caron, Serge Morand, and Michel de Garine-Wichatitsky

Abstract Emerging infectious diseases' hotspots have been identified as multi-host and multi-pathogen systems often characterized in tropical ecosystems by an extensive wildlife/domestic/human interface. The pathogen communities shared by the wild and domestic populations at this interface reflect the historical epidemiological interactions between them. In a research framework using recent community ecology, evolutionary biology and molecular biology advances, this information can be used to identify potential pathways for future pathogen spill-over initiating the emergence process. In other words, an understanding of the mechanisms of pathogen transmission in a specific ecosystem can provide an interaction network between host populations defined by nodes and edges and characterized by the frequency, intensity and direction of the interactions with a direct input for targeted disease surveillance.

14.1 Introduction

The incidence of Emerging Infectious Diseases (EIDs) in human and domestic species has increased in the last decades (Cleaveland et al. 2001; Taylor et al. 2001; Jones et al. 2008; Woolhouse 2008). Zoonoses constitute 60.3% of human EIDs

A. Caron (✉) • M. de Garine-Wichatitsky
Cirad AGIRs, Harare, Zimbabwe

Cirad AGIRs, F-34398 Montpellier, France
e-mail: alexandre.caron@cirad.fr; michel.de_garine-wichatitsky@cirad.fr

S. Morand
Institut des Sciences de l'Evolution, CNRS-IRD-UM2, CC65, Université Montpellier 2, F-34095 Montpellier, France

Cirad AGIRs, F-34398 Montpellier, France
e-mail: serge.morand@univ-montp2.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_14,

of which 71.8% originate in a wildlife source with intermediate animal hosts often necessary (Jones et al. 2008). Following (Haydon et al. 2002), we define target species as those species receiving the bulk of sanitary surveillance: human and domestic species as well as a few wildlife flagship species. The sum of these target species represents a tiny fraction of biodiversity. The bias induced by this focused surveillance hides the majority of EIDs events in the non-target species.

We have little insight about what is happening "in the wild", where statistically most of pathogen evolution and transmission processes are underway. Our knowledge about EIDs processes is therefore limited and skewed in favour of a non-random sample of those occurring. Describing these processes and their common properties could benefit the surveillance and control of EID in target but also in non-target species, supporting global objectives of public and animal health and conservation as presented in the "One World, One Health" concept (Karesh and Deem 2000; Osofsky et al. 2008; Gibbs and Anderson 2009).

The origin of this increase in EIDs is complex. A relation with disturbed ecological processes is often assumed (Daszak et al. 2001). Massive changes in organism distribution and relations induced by global trends, such as increased anthropogenic footprint and climate change have triggered new host, pathogen and environment interactions. These new ecological interactions are not randomly geographically located and regions with a higher risk of emergence can be identified by plotting known emergence events on a map.

According to Jones et al. (2008), hotspots of disease emergence are characterised by:

1. high densities in human and animal populations in systems under intensive health surveillance
2. the wildlife/livestock/human interface in tropical ecosystems.

Focusing on wildlife-linked EIDs, the integration of these two characteristics leads to a description of EIDs events as a two steps process:

1. emergence of the pathogen *sensu stricto* (defined as the interspecies spill-over from a non-target to a target species) (steps 1 and 2 in Childs et al. (2007)
2. amplification of the epidemic phase with higher host availability (provided by higher densities).

Recent EIDs can be classified accordingly: Ebola in humans, originating from a potential bat reservoir has reached step 1 but not step 2 (Leroy et al. 2005); HPAI H5N1 in poultry with a wild bird origin has reached step 1 and 2 (Webster and Govorkova 2006); SARS in humans, originating in bats, as reached step 1 but had a limited step 2 in 2003 (Wang and Eaton 2007). In tropical and sub-tropical ecosystems, the degree of wildlife/livestock/human interface and biodiversity are high and offer numerous potential events of spill-over. However, low animal and human health surveillance decrease the detection probability. In ecosystems with artificially high host densities, less inter-species transmission events occur because of the

physical and sanitary protection of these systems. Moreover, efficient surveillance systems in target species increase the detection probability of the emergence. This analysis could explain the global EID patterns observed.

In order to improve the efficiency and reduce the cost of health interventions, research should be implemented on step 1 in order to understand, predict and control emergence processes and to prevent amplification events (step 2) or prepare the health sector to control them (Barclay 2008). Step 1 requires focusing on the wildlife/livestock/human interface, which necessitates ecological, epidemiological as well as socio-economical information. This interface is characterised by: (a) multi-host systems; (b) managed and unmanaged animal populations ("managed" referring to the state of animal population when influenced by voluntary human intervention); (c) natural and modified habitats; (d) remote areas, where basic commodities are often lacking; (e) unknown socio-economical system in marginalised human communities. The lack of baseline information which characterises these interfaces combined with the above particularities requires the development of theoretical and technological tools adapted to the study of EIDs in their hotspots (Morgan et al. 2006). However, studying EIDs in selected ecosystems raises any of the following challenges: (a) looking for a still unknown pathogen; (b) arriving after the emergence process, when one does not know if the spill-over process still occurs; (c) looking for a pathogen which is not yet in the study site. In order to overcome this dilemma, a shift from a pathogen-centred to a process-centred approach is necessary. The process at stake is the pathogen spill-over in multi-hosts system at the wildlife/livestock/human interface.

We define epidemiological interaction (EI) as any ecological interaction resulting in the transmission of pathogen between two host populations (Caron et al. 2010). This definition of EI is developed in the next sections of this chapter. We then present a research framework, process-orientated, using host and/or pathogen data for the inference of emergence risk in a given ecosystem. We describe a method to build EI network and present, using an example, how this network can be used to identify host populations or EI at risk for pathogen emergence.

14.2 Estimating Transmission Rate for Pathogens Shared Between Host Populations

Standard approaches to study pathogen transmission rate in a single host population are pathogen-centred (Bordes and Morand 2009). In one host population, prevalence and incidence of the pathogen measured by serological or viral detection techniques can be used to calculate the transmission rate. Common index of this transmission rate are the R0 and β index and the force of infection (McCallum et al. 2001; see Real and Biek 2007 for a recent discussion of these parameters for wildlife zoonosis). These parameters are defined for a given pathogen in a defined host population. The data needed to estimate these parameters are the contact rate between hosts, the

transmission probability resulting from these contacts and the infectiveness of the pathogen in the target host or the probability that such a contact occurs between an infected and a susceptible host (McCallum et al. 2001). Models of a shared pathogen between two host populations have been developed (see Tompkins et al. 2002 for some examples).

In an experimental situation using a domestic species, these parameters are accessible. *A posteriori*, after an outbreak, these parameters can be estimated if the relevant data has been timely collected. This data concerns repeated data collection on a pre-determined sample of the animal population and the setting-up of the appropriate environment in which the samples will be adequately stored until laboratory testing. Applied to the wildlife/livestock interface, estimating these parameters is difficult in free-ranging species (Morgan et al. 2006). Capture of wild species is often expensive, implemented in extreme conditions not suitable for sample conservation and unrealistic when the activity needs to be repeated in time on the same animals. Real and Biek (2007) suggest a possible framework using telemetry (radio and satellite) on wild species combined with sampling survey on wild and domestic species. However they identify limits such as the underestimation of the quantity of contacts if the entire wild population is not marked. Richomme et al. (2006) estimated by direct observations the contact rate and exposure between a domestic ungulate and a wild mountain ungulate and discuss additional limits of such data for the inference of EIs because of the nocturnal behaviour of wild species which cannot be apprehended. The necessity to include in the study the respective sensibility of each species to the pathogen requires also an extrapolation from available data on closely-related species. The role of modelling has been and will be crucial in the estimation of the behaviour of transmission parameters and their interpretation (Lloyd-Smith et al. 2009). A model allows playing with variables to estimate outcomes impossible to observe in the field. They use the available data to test hypotheses which would be too complex or too costly to test in experimental conditions. Multi-host models have been developed under the form of meta-population pathogen diffusion process in the particular case of multi-host meta-population (Arino et al. 2005). Recently, transmission models of multi-strains with differential transmission pathways (with emphasis on the role of the environment) paved the way for multi-pathogen models (Roche and Rohani 2010). However, the integration of multi-host and multi-pathogen models has yet to be done. This step will be necessary in order to encompass the full complexity of the ecology of infectious disease transmission at the wildlife/domestic interface.

The application of these different methods to the case of an unknown pathogen before its emergence in the ecosystem raises new issues. Here, the objective is to identify an unknown emerging pathogen to reach the target population. The emergence event *sensu stricto* that we are trying to detect is the spill-over of an unknown pathogen from an unknown source population to a known naive population. Which non-target host should be studied? For which pathogen should we test it? In this context where no target pathogen is identified, a pathogen-centred approach cannot be implemented.

14.3 EI Network and Selection of Host and Pathogen Community to Predict Pathogen Emergence

The human species, a domestic species (e.g., livestock) or a flagship wild species (e.g., mountain gorilla) can be the target species under study. The unknown pathogen emerging in a given ecosystem will be transmitted to the target species by an unknown non-target species, through direct and/or indirect contacts (Fig. 14.1).

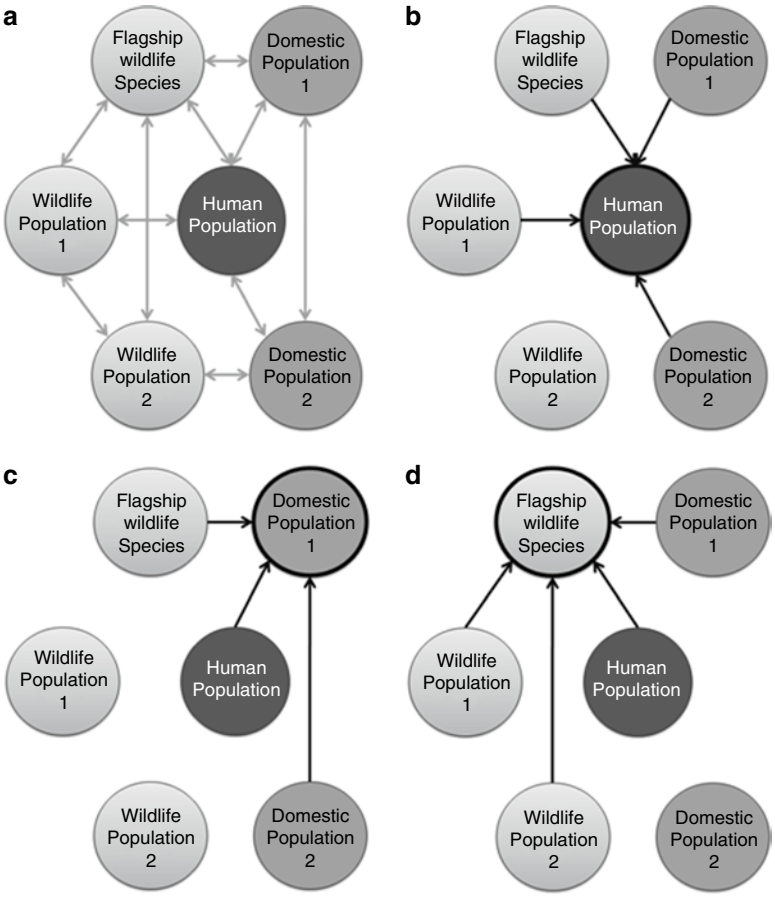

Fig. 14.1 (**a**) Theoretical contact network including wild, domestic and human populations; in the remaining figures, the same network when the human population (**b**), a domestic population (**c**) or a flagship wildlife species (**d**) is considered as the target species. Depending on the target population selected, the contact network potentially leading to the spill-over of a pathogen from non-target populations changes. For example, wildlife population 2 does not play a role in (**b**) and (**c**) whereas it is considered as a potential source of pathogen in case (**d**). Relevant epidemiological interactions in the system change in case (**b**), (**c**) and (**d**)

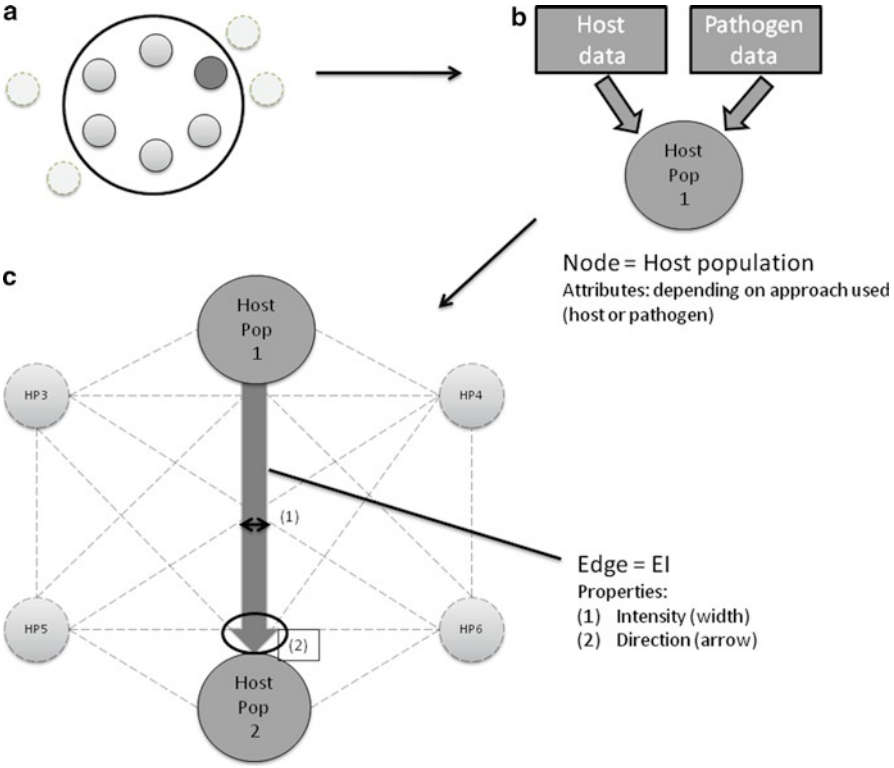

Fig. 14.2 Three-step process to build a EI network and its schematic representation: (**a**) Six Host populations (HP; target – in dark grey- and non-target) are selected in the ecosystem; (**b**) Characterisation of the attributes of the nodes of each HP based on host or pathogen data; (**c**) Zoom on 2 nodes which properties can be decided according to available information (see Box 14.1) and one edge characterised by its 2 properties: (*1*) Intensity represented by the width of the edge; (*2*) Direction represented by an arrow which could uni- or bidirectional

The main assumption of the research framework is that the emerging pathogen will use contacts between hosts which have already been used by other pathogens. The higher the intensity of EI, the higher the probability of the emerging pathogen to use it. If the EI network can be built between different host populations, prediction on the likelihood of emergence through a particular EI can be made.

Network Analysis (NA) developed primarily in the context of social sciences is increasingly used in health ecology (Luke and Harris 2007). It has been applied to explore relations between cattle movement data (Heath et al. 2008) or human populations patterns (Bansal et al. 2007) and disease spread. A network is a set of different entities defined by "nodes" linked by "edges". Nodes and edges have several properties (Fig. 14.2). Nodes are host populations, and edges represent EIs. Nodes can be linked with characteristics of the host population

(ecological or epidemiological) defined by the type of data used to build the network. Edges have two properties:

1. intensity, defined by the number of contacts between host populations, or the proportion of pathogens shared between the two host populations; this property can be graphically represented by the width of the link between 2 nodes
2. directionality, uni- or bidirectional depending on the transmission possibilities between the two host populations; an arrow at the extremity of the edge represents this property.

If interested in the temporal variation of EIs, different networks can be built or different colour for different temporal windows can be used, based on temporal series. Once nodes and edges have been characterised given their attributes, the network illustrates the principal edges-EIs between the target species and non-target species.

The EI network can be build using two different sources of data: host data or pathogen data. The host data consists in estimates of direct and indirect contacts between host populations. The pathogen data consists in the community of pathogens shared by different host populations. In the former case, the host contacts will be used as estimation of future pathogen spill-over and this approach can be called *a priori*. In the later case, pathogen transmissions that have already occurred will be used to estimate the potential pathways for future pathogen spill-over and the approach is in that case *a posteriori*.

The first step in the building of the EI network is to select which host and which pathogen are included in the study. The non-target host populations represent a range of wild and domestic species living in the same environment as the target species, and potentially acting as candidate source of pathogen. As it is usually impossible to obtain relevant data for all wild species in an area for practical and financial reasons, the selection of potential hosts can be prioritised, using available information:

1. Phylogenetic proximity (e.g. apes most closely related to human or wild bovid species most related to livestock species) is a known criteria to increase the chance of pathogen spill-over between two host species (Davies and Pedersen 2008).
2. Ecological knowledge on host species present in the study site.
3. Epidemiological information can orientate as well the selection process.

For example, bats species represent 25% of all mammal species despite being under-studied from an ecological point of view. They can be in contact for behavioural reasons – feeding, roosting – with different host species. They have also been involved in recent human EIDs epidemiology, notably in haemorrhagic fever events (Wang and Eaton 2007). If the target species is the human species, including bats in the study seems relevant if they occur in the ecosystem. Usually, as most studies focus on a specific or a group of pathogens, the selection of host populations is based on the available knowledge on the pathogen's host range. Caron et al. (unpublished) have presented a framework for this selection process based on epidemiological functional groups of hosts. This framework offers a systematic step by step process to consider and select all potential hosts in a given ecosystem. In this selection process, available data about host contact (e.g., questionnaire-based) or

pathogen prevalence (available through national studies) should be used to fine-tune the choice of host populations for the network.

14.4 Estimating Epidemiological Interactions Using Host Data (*A Priori* Approach)

Estimates of contacts between the target population and one or more non-target populations can be used to build the EI network. Host movements and contacts can be monitored using different field techniques adapted to the level of study: individual, population or community.

Few empirical studies have considered contacts at the wildlife/livestock interface. At the individual host level, direct observation or telemetry can detect and quantify these contacts. For example, the development of satellite telemetry and the increase capacity of miniature batteries allow the study of animal movements with a large range of body sizes: up to a few grams for birds (Gaidet et al. 2008) at a very fine time-scale (up to a point every 5 mn) for long periods (a year or more). Recent examples combining telemetry and density or visual observation techniques have successfully determined contact rate of individuals at the wildlife/livestock interface (Legendre et al. 1994; Bohm et al. 2009). The main weakness of these studies is the underestimation of contacts as it can never be assumed that the entire non-target population has been identified. This weakness can be partially controlled by objective or subjective knowledge of the study site but can never be completely addressed.

At the population or community level, estimates of overlaps in habitat used by different host populations or species (Ezenwa 2003; Poulin 2007a) can be utilised as a proxy of interspecies contacts at different seasons. Data collected during road-counts or counts focusing on key resources or habitat (such as waterholes) on both sides of the interface, as well as trapping data (Caley and Hone 2004) provide a quantitative estimate of interspecies interactions which can be used to build an EI network. A recent study attempted to estimate EIs in avian communities using bird count data in wild and domestic sites to predict the risk of Avian Influenza transmission at the wildlife/domestic interface (Caron et al. 2010). Finally, local community questionnaire-based studies can also produce contact rate between host populations, paying particular attention in the design of the questionnaire and treating the information as perceptions and not facts (Brook et al. 2009).

The field of molecular epidemiology has experienced recent and major developments (Gupta et al. 2009). The amount of genetic data on host and pathogen species is produced at an exponential rate (Holmes 2007). Recent years have seen the development of powerful molecular tools to characterize specific pathogens and hosts. This host information supports the characterisation of the genetic distance between and among host species or populations (Nieberding and Olivieri 2007). Host population dynamics estimated by gene flow will be a particular case of population-level study to estimate contacts (see in this book Chevillon et al. 2011; Chapitre de ce livre 2011). Molecular tools in this field have been developed and can give indications as fine as the parental relationship between individuals. This

genetic data can provide fine information to build an EI network between populations of the same species, giving information on the intensity (or width) of edges as well as the direction of edges.

Most of host contact studies to estimate probabilities of pathogen transmission have been developed in the context of a specific pathogen. Our objective is to use the same tools to qualify and/or quantify EI with no *a priori* on the pathogen under study. As presented in Fig. 14.2, after the first step of host selection, each node, representing a host population or species will be attributed characteristics: species and/or population information; and the number of host of the network in contact with the focal species/population (through habitat, resource overlap or direct contact). Edge intensity will be proportional to the estimation of the contact between the two hosts. Edge direction is usually not indicated except in particular circumstances. For example, species A only transiting in a particular habitat will not be infected in this habitat while species B feeding in this habitat could be infected through environmental contamination. Finally, successive EI networks can be built in order to account for temporal variability of contacts (e.g., seasons). The EI network will have to be interpreted with the same limits as the initial contact data: e.g., contacts between populations will be under-estimated.

14.5 Estimating Epidemiological Interactions Using the Pathogen Level (*A Posteriori* Approach)

In the previous paragraph, we presented how host movement and contact data can be used to build an EI network. In the following paragraph, we will show how epidemiological data can also be used to build EI networks using an *a posteriori*, allowing the most likely pathways for the future emergence of new pathogens.

For each pathogen infecting more than one host population, we can estimate EIs; accumulating data on an array of different pathogens will strengthen the network. In other words, gathering epidemiological data on numerous pathogens in an ecosystem in different wild and domestic populations can structure the EI network between these populations and predict to a certain extent (discussed below) the behaviour of an outsider pathogen. The type of data referred to here as already been collected in various studies, albeit not used for this purpose. Jolles et al. (2008) collected data on micro- and macroparasite species in buffaloes in South Africa to investigate potential ecological interactions (competition, synergies) between pathogens in a single host. In another study, Ezenwa 2003 looked at macroparasite richness in sympatric wild and domestic ungulates, and analysed their variations according to habitat overlap between host species. Other studies dealing with the health status of some endangered species, a parameter increasingly investigated in species conservation, can also produce relevant information on free-ranging species (Philippa et al. 2008).

Parasite community ecology is a subset of community ecology focusing on the distribution of parasites between host populations. Large datasets of parasites (mostly macroparasites) in multi-hosts systems have been collected in the course of these studies (see Poulin 2007b). Several indices have been used to compare parasite

communities between host populations (Poulin 2003, 2010). The Jaccard index (Jaccard 1912) uses presence-absence data of parasites in different host populations to give an indicator of the similarity of this community. Other indices, more sophisticated, use prevalence data to compare parasite community (Boyle et al. 1990; Poulin 2007b). In the context of host populations from the same ecosystem, the geographic distance between the host populations is controlled (geographic distances usually looked at in parasite community ecology are of the order of hundreds or thousands of kilometres). If the phylogenetic distance between host populations can be controlled, the similarities measured can estimate the EI between these host populations. These indices can therefore be used to build EI network as the value of the index between two parasite communities shared by two host populations (their similarity). Analytical tools to control for phylogentic distances between hosts are already used and new ones are under development that incorporate sophisticated statistical analyses (Adams 2008; Hadfield and Nakagawa 2010).

History of contacts between host populations is another variable to be considered (see Sect. 14.6). The type of the epidemiological data that one can gather varies depending on the pathogens targeted and the investment of the scientific and private community into the development of specialised diagnostic tools. Presence-absence data are the simplest data that one can gather on pathogen epidemiology in a singular host, and is often the only available information from the literature. Prevalence data (direct through pathogen detection or indirect through antibody detection) provides the percentage of hosts in each population which have been or are in contact with the pathogens. As mention above, this type of data can be included in the calculation of similarity indices borrowed from community ecology (Boyle et al. 1990; Poulin 2007b).

As presented in the previous section, the molecular epidemiology revolution has changed the field of parasite ecology and new genetic information is providing the ground for major advances in pathogen research, evolutionary ecology and population dynamics of pathogens. The characterisation of HIV strains from different human and great apes populations or even from different human individuals has brought important development on our knowledge about the origin and the spread of this pathogen across the world (Heeney et al. 2006; Cohen 2007; Gilbert et al. 2007). Another example is the abundant recent literature on phylogenetic analyses of HPAI H5N1 strains across the globe linking animal and human outbreaks (see Wang et al. 2008; Cattoli et al. 2009). Other important pathogens have benefited from these technological advances (for some examples on multi-host utilisation see Bastos et al. 2003; Vosloo et al. 2006; Foster et al. 2009). The level of accuracy to detect single nucleotide change in parasite genomes of these molecular tools is increasing and their use at the ecosystem level can pick-up recent transmission events including their direction. Biek et al. 2006 have used the Feline Immunodefficiency virus (FIV) to track the population dynamics of one of its host (*Puma concolor*). Johnson et al. (2010) highlight this point with another angle and suggest to compare the outcomes of an EIDs outbreak with "closely related pathogens in different but related host species (…)". The evolution of parasites is often faster than the evolution of their hosts (Nieberding and Olivieri 2007). Recent host

population dynamics not detected in the host genetic material can be captured by the faster genetic evolution of its parasites. Applied to our context, a phylogenetic tree of the same parasites detected in different host populations in the same ecosystem can reveal connexions between these host populations and give an estimate of the EIs between these host populations for this specific pathogen. The molecular tools can be used to characterise EIs. It is likely that in the coming years additional genetic data on parasites and new analytical methods will provide more power in the estimation of EIs between host populations. Transmission pathways between host populations and EI network will be strengthened by these advances (Chevillon et al. 2011).

Replicating the multi-step process developed in Fig. 14.2 and in the previous section, once the host selection has been implemented in the ecosystem, each node representing a host population can be characterised by the total number of parasite species harboured (nodes property represented by its size) and other attributes (e.g. in Box 14.1). To estimate the intensity of the edge (its width), we use the value of the index calculating the proportion of the parasites shared between two host populations/species. In very rare occasions where knowledge about the epidemiology of some parasites exists, the direction of the EI could be known: e.g., when a species known to be susceptible but not capable of maintaining a parasite is infected by a known reservoir of the parasite. Molecular data can potentially inform EIs on both their properties: intensity, as molecular data can be used as prevalence data (e.g. prevalence for different strains of parasites); direction, the evolution of a parasite strain between two host populations can be tracked back and the donor population can be distinguished from the receptor population. Finally, if time series data is available different network can estimate the variation of EIs across seasons (time series used to infer transmission dynamics between hosts see Begon et al. 1999).

So far, the use of these epidemiological and molecular tools has been limited to one target pathogen and only a few articles have referred to the extension of their use to several pathogens, to host population dynamics, or to contacts between hosts (Poss et al. 2002). At an ecosystem level, it is the combined use of presence-absence, prevalence and molecular data on multiple pathogens which will define EI networks. The integration of different type of data for different parasite species will be a challenge in the framework of this approach. Prevalence data will give more fine-scale details about the shared community of pathogens than presence/absence data as for each host population, the estimated percentage of individuals in the population infected by the parasite will be taken into account in the index. Molecular data can give even more details about how long ago the parasite strain was transmitted from one host to the other. To our knowledge, the only possibility to integrate this data in one network is to weight each type of data (e.g., giving more weight to molecular, then prevalence and finally presence/absence data). The more data gathered on various pathogens, the more detailed the hypotheses on future EIDs epidemiological pathways.

For example, de Garine-Wichatitsky et al. (2010) have recently detected the first case of bovine tuberculosis (bTB) in African buffaloes in a Zimbabwean national park and discussed the outcomes of this emergence at the country level. How will the pathogen behave in this ecosystem? Will it spill-over to communal cattle populations? To other wildlife species? To human populations? Data on circulating diseases

(zoonoses and others) in this area (e.g. foot-and-mouth disease, brucellosis, rift valley fever etc.) can bring information on the structure of the EI network. The network and other epidemiological knowledge about parasites can be used to prioritize the nodes or edges at risk. One can consider interventions on the nodes (host) of the network for surveillance and/or control but also on the edges (EIs) to "break" the transmission pathways and limit the potential for emergence. If not detailed here, the relation between an EI network and a spatially explicit approach can be done.

Box 14.1 Epidemiological Interaction Network for 14 Rodent Species and the Human Species

In this example, the human species is our target species and we explore the EIs between the human species and several rodent species present in particular ecosystems of Southeast Asia represented by different habitats (dry and irrigated agricultural areas, forests, and villages).
From the literature (Chaisiri et al. 2010) (Herbreteau et al. unpublished), we obtained presence-absence data on 14 rodent species. Information about 34 macroparasite species and 8 microparasite species were collected for these 14 rodent species and susceptibility to these parasites for the human species were obtained (Table 14.1). A 42 parasites*15 hosts matrix was built (and filled with "1" or "0" for occurrence of infection and absence respectively in each host species. This matrix was used to calculate the Jaccard Index (=number of parasite species present in both host populations/sum of parasite species present in each host populations) displayed in Table 14.2. The Jaccard index value

Table 14.1 Host and parasite species used

Target sp.	*Homo sapiens*
Rodent sp.	*Bandicota indicata (Bi), Bandicota savilei (Bs), Verylmys bowersi (Bb), Leopoldamys edwardsi (Le), Maxomys surifer (Ms), Mus caroli (Mc), Niviventer fulvescens (Nf), Rattus andamanensis (Ran), Rattus argentiventer (Rar), Rattus exulans (Re), Rattus losea (Rl), Rattus norvegicus (Rn), Rattus tanezumi (Rta), Rattus tiomanicus (Rti)*
Macroparasite sp.	*Hymenolepis nana, Rodentolepis sp., Taenia sp., Taenia taeniaeformis, Ascaris sp., Gnathostoma malaysiae, Ganguleterakis spumosa, Citellina levini, Syphacia muris, Physaloptera sp., Rictularia sp., Rictularia tani, Gongylonema neoplasticum, Mastophorus muris, Protospiura-Mastophorus sp., Cyclodontostomum purvisi, Strongyloides ratti, Strongyloides sp., Nippostrongylus brasillensis, Nippostrongylus sp., Orientostrongylus tenorai, Echinostoma ilocanum, Echinostoma malayanum, Notocotylus sp., Quinqueseralis quinqueseralis, Gastrodiscoides hominis, Centrocestus sp.*
Microparasite sp.	*Leptospira spp, Ornentia spp, Bartonella spp*, hantavirus, herpes virus, LCM virus, *Trypanosoma spp*, rabies virus.

(continued)

Box 14.1 (continued)

Table 14.2 Jaccard index of shared parasite community for each pair if host poulation

	Nb Para	Bi	Bs	Bb	Le	Ms	Mc	Nf	Ran	Rar	Re	Rl	Rn	Rta	Rti	Hs
Bi	**16**															
Bs	**9**	0.47														
Bb	**7**	0.15	0.00													
Le	**3**	0.12	0.09	0.11												
Ms	**8**	0.20	0.06	0.15	0.22											
Mc	**2**	0.06	0.00	0.00	0.00	0.25										
Nf	**2**	0.06	0.10	0.00	0.00	0.00	0.00									
Ran	**5**	0.24	0.17	0.09	0.33	0.30	0.40	0.17								
Rar	**17**	0.43	0.24	0.14	0.00	0.09	0.06	0.12	0.10							
Re	**18**	0.48	0.35	0.00	0.11	0.13	0.05	0.05	0.15	0.30						
Rl	**9**	0.19	0.29	0.00	0.00	0.06	0.00	0.22	0.08	0.37	0.23					
Rn	**23**	0.34	0.23	0.03	0.04	0.07	0.04	0.09	0.08	0.33	0.52	0.19				
Rta	**32**	0.45	0.24	0.11	0.06	0.14	0.03	0.06	0.09	0.32	0.47	0.17	0.72			
Rti	**13**	0.26	0.05	0.05	0.00	0.17	0.07	0.07	0.06	0.36	0.24	0.10	0.33	0.36		
Hs	**15**	0.48	0.50	0.05	0.06	0.21	0.06	0.06	0.18	0.28	0.43	0.20	0.41	0.42	0.17	

Nb Para: the number of parasites detected by host species

(continued)

Box 14.1 (continued)

varies therefore between "0" for no parasite species shared and "1" for all parasite species shared.

We used the Jaccard index as a proxy of the epidemiological interactions (EIs) between each host population and built the corresponding EI network (Fig. 14.3).

The analysis of this network leads to the following observations:

– Interpreting the size of the nodes, the three rodent species with the highest parasite diversity are occurring in human settlements. The three rodent species with the lowest parasite diversity occur in primary or secondary forests and dry agricultural land.

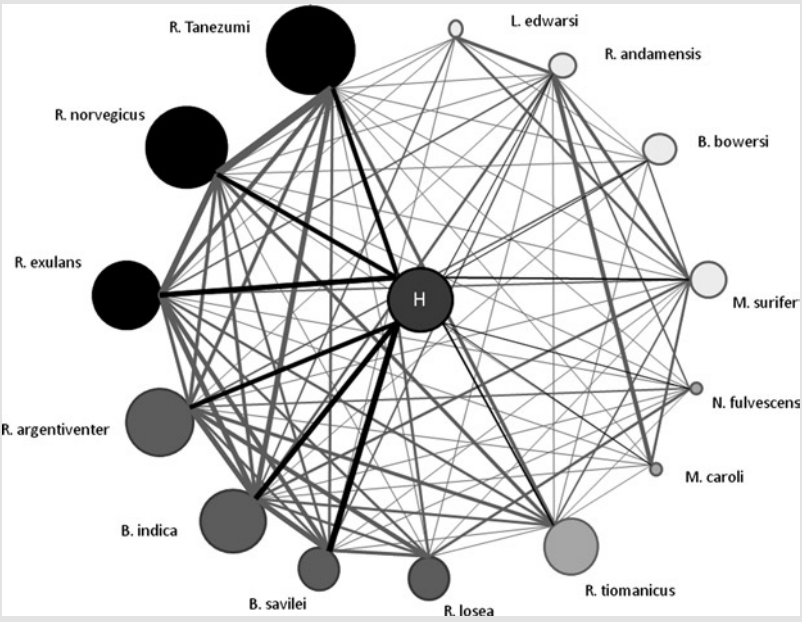

Fig. 14.3 Epidemiological Interaction Network for 14 rodent species and the human species in the Southeast Asian ecosystems based on presence-absence data for 34 macroparasite species and 8 microparasite species. Each node represents a host species, the size of the node is proportional to the number of parasite species harboured by the host and the greyed level of the circle represents the habitat in which the host species is mostly found (except for human): from pale grey for primary forest, rice fields, modified forest, dry agricultural areas to black for human settlements. Each edge between two nodes represents the shared parasite community and its width is proportional to the Jaccard index. We placed the human species in the centre of the figure and its edges in black for visual comfort

(continued)

Box 14.1 (continued)

- The size of the human species node indicates that we share 15 parasite species with rodent species studied here.
- Interpreting the width of the edges at the network level, there is a higher density of large-width edges on the left of the network, indicating that rodent species in human settlements and rice-fields share a higher proportion of their parasite diversity than rodent species in the remaining habitats.
- Interpreting the width of the edges concerning the human species, Bi and Bs have the highest Jaccard index values (0.48 and 0.5 respectively), followed by Re, Rta and Rn (0.43, 0.42, 0.41 respectively).
- The nodes of Bi, Ran, Rn and Rta have the maximum number of edges (n = 14) possible in this network. They all occur in human settlements, rice fields except for Ran occurring in primary forest. The nodes of Bb, Le and Mc have the lowest number of edges in the network (n = 9) and they all belong to primary and secondary forest or dry agricultural areas.
- The node of the human species has 13 edges close to the maximum of 14.

The method of calculus of the index needs to be kept in mind: the observation that Bi and Bs share more parasites with the human species than Rta and Rn is wrong. The human species shares more parasite species with Rn and Rta than with Bi and Bs (11, 14, 10, 8 respectively). The difference is due to the high parasite species richness of Rn and Rta. Other indices can be used to address this kind of issue but no index is perfect. The highest intensities of EI are found in rodent species present in heavily disturbed habitat (human settlement and rice fields) and the highest intensities of EI linking the human species with rodent species are also found with rodent species living in these same habitats. This observation suggests that humans are more at risk of contracting new pathogens from rodents present in the human settlement compared to rodents present in primary forest. The availability and patchiness of resources in human modified environment could explain the higher EI intensity in rodent species living in these environments.

This first network can orientate surveillance protocols towards the most interesting host species to be included in order to answer the question at stake: if the question is the probability of EID from rodent host in this ecosystem, the surveillance protocol will target species living in the human settlements (and a ranking can be done on this species) and in the rice fields with maybe Rti being an interesting sentinel species to look at as a bridge species between pristine and modified environment. To our knowledge, this species is never mentioned as a potential source of infectious disease or as a sentinel in the literature. If one is more interested in the potential emergence of pathogens in rodent species, Ran which shares parasite species with all the other rodent species and humans and lives mostly in primary forests should be considered as a potential source.

14.6 Discussion

The discussion will mainly address the pathogen approach as the host approach is at advance stage of development in the literature.

From an epidemiological perspective, usually investigating one epidemiological cycle at a time, the EI network approach that we present makes no sense. The epidemiology of each disease is different and one cannot draw inferences on a pathogen using data from another. However, from an ecological perspective, the line of thoughts is logical even if challenging. There is not an infinite opportunities for a pathogen to be transmitted between two host populations and new pathogens will use routes already used by other pathogens with a higher probability. Identifying common pathways between host populations will therefore increase the knowledge about future probable emergence pathways in this system. This pathogen approach does not refer to a specific pathogen and focuses on the hosts' direct and indirect contacts but only the ones resulting in pathogen transmission. The utility of such an approach is obvious in the surveillance of EIDs in a hotspot: with an EI network, one can assess the probability of future spill-over processes between specific populations and target surveillance and/or control. In a limited-resource environment (true everywhere but even more in most EID hotspots), EI networks will point at most probable transmission pathways between host populations which can be acted upon to reduce the probability of transmission/emergence.

Box 14.1 gives an example of an EI network built using presence-absence data of macro- and micro- parasites in 14 rodent species and the human species. The data has been gathered from the literature (see Chaisiri et al. 2010). Using a simple method to describe the parasite communities shared by each pair of host populations, the EI network illustrates the intensity of each edge proportional to its width, the parasite species richness per host and the main habitat where each host occurs. Eco-epidemiological information such as which rodent host shares a high proportion of its parasite species with humans are visually explicit and new hypotheses about key-rodent species are presented. These observations suggest that the EI network is representative of the level of contacts between the human species and the rodent species. Finally, combining this (crude) epidemiological and ecological information in a single analysis provide more information about the system than a simple juxtaposition of single parasite information.

In several practical situations, scientists and managers will be concerned by the possible emergence of a specific pathogen in a given ecosystem and for a specific target species. In that case, they should target closely related pathogens or some with similar behaviour (transmission pathways, host range) in order to delineate a more specific EI network. This step can be done by creating a sub-network with the selected pathogens or by weighting the influence of some pathogens in the network. Other *a priori* conditions could be a research question targeting a type of emerging pathogen to look for (e.g., haemorrhagic fevers). For example, several authors have argued that future EIDs will concern in majority RNA virus-type pathogens, due to inherent viral characteristics such as high mutation rate

(Poss et al. 2002; Holmes and Rambaut 2004; Cleaveland et al. 2007). The design of a research framework focusing on emerging RNA viruses will select pathogens with similar mode of transmission (other RNA viruses for example) which should use the same EIs. This could be done by focusing the EI network with information from parasites phylogenetically related to the potentially emerging parasite of interest (e.g., separating viruses, bacteria, macroparasites, etc.) and/or sharing the same mode(s) of transmission. The later seems to bear more power than the former, but this will need to be tested. This approach is process orientated and the modes of transmission of parasite will be crucial in determining the transmission pathways between host populations. Groups of parasite with direct, environmental or vector-born transmission could *a priori* produce different EI networks. Three outcomes could emerge from empirical data:

1. each EI network is specific for each parasite species and the EI network approach is somehow useless
2. they are functional groups of parasite sharing a similar EI network (e.g., based on their modes of transmission)
3. EI network are general and produce a transmission framework for all parasite species given that the network is fed with data from enough parasite species.

In the case study of Box 14.1, sub-network could be explored comparing macro- and microparasites EI network or regrouping parasite by modes of transmission and comparing the corresponding sub-network.

Another argument in favour of the inclusion of all types of parasites in the building of EI network using the *a posteriori* approach is that there is no independence between the parasite data from various hosts (Jolles et al. 2006). Each parasite species is in struggle with the host immunity system and in ecological interaction (direct e.g., competition, synergy or indirect through the host immune system) with the other parasites. The community of parasites in the ecosystem will have its own ecological interactions defined by host and parasite presence and densities (Atmar and Patterson 1993; Booth and Dunne 2004). Despite some interesting recent results (Graham 2008; Jolles et al. 2008), co-infections of parasites in host populations have not been much explored. The patterns of co- occurrence or exclusion between parasites in hosts are *de facto* included in EI networks and will not create issues for their interpretation. Developing EI network will provide more information on parasite interaction in hosts.

Host populations in the same ecosystem for long periods of time should share more parasites than host populations which have recently come into contact. Here, "recently" refers to a period of time during which host populations did not have time to exchange most of their parasite species. Evolutionary processes should be close to equilibrium in co-evolved host-pathogen interaction (e.g. for Low Pathogenic Influenza viruses and waterfowl, Webster et al. 1992) and the EI network identified quite robust. As time since hosts have been in contact increases, potential transmission events with low probability will statistically increase. In the EID hotspots considered, the time since first contact between a target and a non-target population should be small: human and great apes due to hunting or tourism; livestock

and wild ungulates in pristine areas recently colonised; intensive poultry production units recently installed in proximity of wetlands used by wild birds. In EIDs hotspots, by definition, (human-induced) changes disturb ecological interactions in communities. In this context, the EI identified could be instable (Altizer et al. 2003; Thomson 2005) and prone to change leading to an evolving EI network. In this context, the *a priori* approach, using host movements and contacts could be used. The *a posteriori* approach will be more difficult to implement in extremely recent interface. However, a few years should be enough for the shared community of parasites to reach a relative equilibrium.

More information could be added in the EI network presented in Box 14.1. First, some of the parasites used, mainly microparasites such as *Leptospira* or Hantaviruses, have benefited from molecular tools able to track minute changes in their genetic load. As mentioned already microparasites are fast evolving organisms compared to their hosts (Poss et al. 2002; Holmes and Rambaut 2004). This can be useful in tracking the origin of infection between different populations, capturing the history of spill-over process at a fine scale. A phylogenetic tree of the parasites strains detected in different host populations could bring information on the directionality of some EI and indicates rodent species as source or reservoir for the rest of host community (for an example see Cottam et al. 2008). Secondly, data from the host population, as presented previously, such as home range or telemetry data could be included to incorporate potential direct and indirect contact between host populations and related to EI identified in the pathogen approach. This would combine the host and pathogen approaches.

14.7 Conclusion

After presenting how the use of host data can shed light on transmission pathways between host populations, we developed a multi-pathogen approach, process-centred, to infer epidemiological interaction at the wildlife/livestock/human interface. Theoretical and technical tools for this approach have already been developed but they are used with a different angle here. Epidemiological data on various pathogens are integrated in a network to predict the behaviour of EIDs, in particular before the emergence event. The need for this approach comes from an empirical point of view, from experiences in areas of limited resources (financial and available data) and tries to answer practical questions of surveillance and control of EIDs where we know that they have a high probability to happen (EID hot spots). The scale of study – the community level – is extremely complex but we suggest that we could benefit from this scale in a resource limited environment.

Albeit empirical, this approach links with the modelling approach by concentrating on processes of transmission. The data produced should feed some models (e.g. Arino and van den Driessche 2006) and maybe push the building of models towards including multi-pathogen data to construct EI networks. Network-based modelling can be an entry point (Legendre et al. 1994) with the potential to include

social data, facilitating the inclusion of human/animal transmission processes (Legendre et al. 1994; Van Kerkhove et al. 2009).

The context of EIDs in developing countries is an environment where usually little epidemiological data is available except for key diseases, important for government services: if existing data on a pathogen brings light on contacts between two host species, this builds a starting point for the EI network. The gathering of ecological and epidemiological information available for the ecosystem under study, obtained from different sources (e.g., literature, veterinary services, conservation NGOs) can be included in a preliminary EI network which could identify the first surveillance priorities to protect the target species following for example a risk analysis process. The inclusion of the epidemiology of several pathogens in the analysis should explore the results of shared prevalence or strains in relation to the specific modes of transmission. If molecular data exists for a pathogen, the comparison of this data across host populations can say a lot about the history of contacts between these hosts and about the intensity and the frequency of contacts (Real et al. 2005; Biek et al. 2006).

From a more practical point of view, one could argue that such an approach would be too costly to be funded mainly due to the multiplicity of diagnostic tests which can peak quickly when gene sequencing is needed. The cost of sampling wildlife species in remote areas is high. Specialised team working on particular diseases will be more than willing to collaborate and contribute in terms of laboratory cost in order to access such rare samples. This requires prior communication and agreement with interested teams and logistical arrangements for the right samples to be collected and delivered to laboratories. The accumulation of such collaborations will increase the multi-pathogen data and strengthen the EI network. Multidisciplinarity inter- and intra- research team is a prerequisite for this approach.

Acknowledgments This study was part of a French ANR Biodiversity ANR 07 BDIV 012, project CERoPath, "Community Ecology of Rodents and their Pathogens in a changing environment". The authors are also grateful to Kittipong Chaisiri and Vincent Herbreteau.

References

Adams DC (2008) Phylogenetic meta-analysis. Evolution 62:567–572

Altizer S, Harvell D, Friedle E (2003) Rapid evolutionary dynamics and disease threats to biodiversity. Trends Ecol Evol 18:589–596

Arino J, van den Driessche P (2006) Disease spread in metapopulations. Fields Inst Commun 48:1–12

Arino J, Davis JR, Hartley D et al (2005) A multi-species epidemic model with spatial dynamics. Math Med Biol 22:129–142

Atmar W, Patterson BD (1993) The measure of order and disorder in the distribution of species in fragmental habitat. Oecologia 96:373–382

Bansal S, Grenfell BT, Meyers LA (2007) When individual behaviour matters: homogeneous and network models in epidemiology. J R Soc Interface 4:879–891

Barclay E (2008) Predicting the next pandemic. Lancet 372:1025–1026

Bastos AD, Anderson EC, Bengis RG et al (2003) Molecular epidemiology of SAT3-type foot-and-mouth disease. Virus Genes 27:283–290

Begon M, Hazel SM, Baxby D et al (1999) Transmission dynamics of a zoonotic pathogen within and between wildlife host species. Proc Roy Soc Lond Ser B 266:1939–1945

Biek R, Alexei J, Drummond J et al (2006) A virus reveals population structure and recent demographic history of its carnivore host. Science 311:538–541

Bohm M, Hutchings MR, White PC (2009) Contact networks in a wildlife-livestock host community: identifying high-risk individuals in the transmission of bovine TB among badgers and cattle. PLoS One 4:e5016

Booth M, Dunne DW (2004) Spatial awareness in parasite immuno-epidemiology. Parasite Immunol 26:499–507

Bordes F, Morand S (2009) Parasite diversity: an overlooked metric of parasite pressures. Oikos 118:801–806

Boyle TP, Smillie GM, Anderson JC et al (1990) A sensitivity analysis of nine diversity and seven similarity indices. Res J Water Pollut C 62:749–762

Brook RK, Kutz SJ, Veitch AM et al (2009) Fostering community-based wildlife health monitoring and research in the Canadian North. EcoHealth 6:266–278

Caley P, Hone J (2004) Disease transmission between and within species, and the implications for disease control. J Appl Ecol 41:94–104

Caron A, de Garine-Wichatitsky M and Morand S (Submitted) Ecology of infectious disease transmission in multi-host systems

Caron A, de Garine-Wichatitsky M, Gaidet N et al (2010) Estimating dynamic risk factors for pathogen transmission using community-level bird census data at the wildlife/domestic interface. Ecol Soc 15:25

Cattoli G, Monne I, Fusaro A et al (2009) Highly pathogenic avian influenza virus subtype H5N1 in Africa: a comprehensive phylogenetic analysis and molecular characterization of isolates. PLoS One 4:e4842

Chaisiri K, Chaeychomsri W, Siruntawineti J et al (2010) Human-dominated habitats and helminth parasitism in Southeast Asian murids. Parasitol Res 107:931–937

Chapitre de ce livre (2011) Molecular epidemiology of disease resistance genes with perspectives for researches on biological invasions and hybrid zones. In: Morand S, Beaudeau F, Cabaret S (eds.) New frontiers in molecular epidimiology. Springer (in press)

Chevillon C, de Meeûs T, McCoy KD (2011) Population genetics and molecular epidemiology of infectious diseases. In: Morand S, Beaudeau F, Cabaret S (eds.) New frontiers in molecular epidimiology. Springer (in press)

Childs JE, Richt JA, Mackenzie JS (2007) Introduction: conceptualizing and partitioning the emergence process of zoonotic viruses from wildlife to humans. In: Childs JE, Mackenzie JS, Richt JA (eds.) Wildlife and emerging zoonotic diseases: the biology, circumstances and consequences of cross-species transmission. Springer, Heidelberg, pp 1–31

Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. Proc R Soc Lond Ser B 356:991–999

Cleaveland S, Haydon DT, Taylor L (2007) Overview of pathogen emergence: which pathogens emerge, when and why? In: Childs JE, Mackenzie JS, Richt JA (eds.) Wildlife and emerging zoonotic diseases: the biology, circumstance and consequences of cross-species transmission. Springer, Heidelberg, pp 85–111

Cohen J (2007) Reconstruction of the origins of the AIDS epidemic from archived HIV isolates. Science 318:731

Cottam EM, Thebaud G, Wadsworth J et al (2008) Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. Proc R Soc Lond Ser B 275:887–895

Daszak P, Cunningham AA, Hyatt AD (2001) Anthropogenic environmental change and the emergence of infectious diseases in wildlife. Acta Trop 78:103–116

Davies TJ, Pedersen AB (2008) Phylogeny and geography predict pathogen community similarity in wild primates and humans. Proc R Soc Lond Ser B 275:1695–1701
de Garine-Wichatitsky M, Caron A, Gomo A et al (2010) Bovine tuberculosis in buffaloes, Southern Africa. Emerg Infect Dis 16:884–885
Ezenwa VO (2003) Habitat overlap and gastrointestinal parasitism in sympatric African bovids. Parasitology 126:379–388
Foster JT, Beckstrom-Sternberg SM, Pearson T et al (2009) Whole-genome-based phylogeny and divergence of the genus Brucella. J Bacteriol 191:2864–2870
Gaidet N, Cattoli G, Hammoumi S et al (2008) Evidence of infection by H5N2 highly pathogenic avian influenza viruses in healthy wild waterfowl. PLoS Pathog 4:e1000127
Gibbs EPJ, Anderson TC (2009) 'One World – One Health' and the global challenge of epidemic diseases of viral etiology. Vet Ital 45:35–44
Gilbert MTP, Rambaut A, Wlasiuk G et al (2007) The emergence of HIV/AIDS in the Americas and beyond. Proc Natl Acad Sci USA 104:18566–18570
Graham AL (2008) Ecological rules governing helminth-microparasite coinfection. Proc Natl Acad Sci USA 105:566–570
Gupta R, Michalski MH, Rijsberman FR (2009) Can an infectious disease genomics project predict and prevent the next pandemic? PLoS Biol 7:e1000219
Hadfield JD, Nakagawa S (2010) General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. J Evol Biol 23:494–508
Haydon DT, Cleaveland S, Taylor LH et al (2002) Identifying reservoirs of infection: a conceptual and practical challenge. Emerg Infect Dis 8:1468–1473
Heath MF, Vernon MC, Webb CR (2008) Construction of networks with intrinsic temporal structure from UK cattle movement data. BMC Vet Res 4:11
Heeney JL, Dalgeish AG, Weiss RA (2006) Origins of HIV and the evolution of resistance to AIDS. Science 313:462–466
Holmes EC (2007) Viral evolution in the genomic age. PLoS Biol 5:e278
Holmes EC, Rambaut A (2004) Viral evolution and the emergence of SARS coronavirus. Proc R Soc Lond Ser B 359:1059–1065
Jaccard P (1912) The distribution of the flora in the alpine zone. New Phytol 11:37–50
Johnson PT, Dobson A, Lafferty KD et al (2010) When parasites become prey: ecological and epidemiological significance of eating parasites. Trends Ecol Evol 25:362–71
Jolles AE, Etienne RS, Olff H (2006) Independent and competing disease risks: implications for host populations in variable environments. Am Nat 167:745–757
Jolles AE, Ezenwa VO, Etienne RS et al (2008) Interactions between macroparasites and microparasites drive infection patterns in free-ranging African buffalo. Ecology 89:2239–2250
Jones KE, Patel NG, Levy MA et al (2008) Global trends in emerging infectious diseases. Nature 451:990–994
Karesh WB, Deem SL (2000) Conservation medicine: a veterinary perspective. Cons Biol 14:336–337
Legendre P, Lapointe F-J, Casgrain P (1994) Modeling brain evolution from behavior: a permutational regression approach. Evolution 48:1487–1499
Leroy EM, Kumulungui B, Pourrut X et al (2005) Fruit bats as reservoirs of Ebola virus. Nature 438:575–576
Lloyd-Smith JO, George D, Pepin KM et al (2009) Epidemic dynamic at the human-animal interface. Science 326:1362–1367
Luke DA, Harris JK (2007) Network analysis in public health: history, methods, and applications. Annu Rev Public Health 28:69–93
McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modelled? Trends Ecol Evol 16:295–300
Morgan ER, Lundervoldb M, Medleyb GF et al (2006) Assessing risks of disease transmission between wildlife and livestock: the Saiga antelope as a case study. Biol Conserv 131:244–254

Nieberding C, Olivieri I (2007) Parasites: proxies for host history and ecology? Trends Ecol Evol 22:156–165
Osofsky SA, Cumming HM, Kock MD (2008) Transboundary management of natural resources and the importance of a "One Health" approach. In: Fearn E (ed.) State of the wild: a global portrait of wildlife, wildlands, and oceans. Island, Washington, DC
Philippa J, Fournier-Chambrillon C, Fournier P et al (2008) Serologic survey for selected viral pathogens in free-ranging endangered European mink (Mustela lutreola) and other mustelids fro the South-Western France. J Wildl Dis 44:791–801
Poss M, Biek R, Rodrigo A (2002) Viruses as evolutionary tools to monitor population dynamics. In: Aguirre AA, Ostfeld RS, Tabor GM et al (eds.) Conservation medicine: ecological health in practice. Oxford University Press, Oxford, pp 118–129
Poulin R (2003) The decay of similarity with geographical distance in parasite communities of vertebrate hosts. J Biogeogr 30:1609–1615
Poulin R (2007a) Are there general laws in parasite ecology? Parasitology 134:763–776
Poulin R (2007b) Evolutionary ecology of parasites, 2nd edn. Princeton University Press, Princeton
Poulin R (2010) Decay of similarity with host phylogenetic distance in parasite faunas. Parasitology 137:733–741
Real LA, Biek R (2007) Infectious disease modeling and the dynamics of transmission. In: Childs JE, Mackenzie JS, Richt JA (eds.) Wildlife and emerging zoonotic diseases: the biology, circumstance and consequences of cross-species transmission. Springer, Heidelberg, pp 33–50
Real LA, Henderson JC, Biek R et al (2005) Unifying the spatial population dynamics and molecular evolution of epidemic rabies virus. Proc Natl Acad Sci USA 102:12107–12111
Richomme C, Gauthier D, Fromont E (2006) Contact rates and exposure to inter-species disease transmission in mountain ungulates. Epidemiol Infect 134:21–30
Roche B, Rohani P (2010) Environmental transmission scrambles coexistence patterns of avian influenza viruses. Epidemics 2:92–98
Taylor LH, Latham SM, Woolhouse MEJ (2001) Risk factors for human disease emergence. Proc R Soc Lond Ser B 356:983–989
Thomson JN (2005) Coevolution: the geographic mosaic of coevolutionary arms races. Curr Biol 15:992–994
Tompkins DM, Dobson AP, Arneberg P et al (2002) Parasites and host population dynamics. In: Hudson PJ, Rizzoli A, Grenfell BT et al (eds.) The ecology of wildlife diseases. Oxford University Press, Oxford, pp 45–62
Van Kerkhove MD, Vong S, Guitian J et al (2009) Poultry movement networks in Cambodia: implications for surveillance and control of highly pathogenic avian influenza (HPAI/H5N1). Vaccine 27:6345–6352
Vosloo W, Bastos AD, Boshoff CI (2006) Retrospective genetic analysis of SAT-1 type foot-and-mouth disease outbreaks in southern Africa. Arch Virol 151:285–298
Wang L-F, Eaton BT (2007) Bats, civets and the emergence of SARS. In: Childs JE, Mackenzie JS, Richt JA (eds.) Wildlife and emerging zoonotic diseases: the biology, circumstance and consequences of cross-species transmission. Springer, Heidelberg, pp 325–344
Wang G, Zhang D, Li L et al (2008) H5N1 avian influenza re-emergence of Lake Qinghai: phylogenetic and antigenic analyses of the newly isolated viruses and roles of migratory birds in virus circulation. J Gen Virol 89:697–702
Webster RG, Govorkova EA (2006) H5N1 influenza – continuing evolution and spread. N Engl J Med 21:2174–2177
Webster RG, Bean WJ, Gorman OT et al (1992) Evolution and ecology of influenza A viruses. Microbiol Rev 56:152–179
Woolhouse ME (2008) Emerging diseases go global. Nature 451:898–899

Chapter 15
Conclusion

Serge Morand, Jacques Cabaret, and François Beaudeau

The contributors of this book have explored various facets of molecular epidemiology with major references to population genetics, ecology and evolution.

Molecular epidemiology is now confronted to new challenges in relation to the impacts of global changes on infectious diseases as summarized by Vourc'h et al. (Chap. 13), that favor the spread of drug resistance in parasites of livestock as shown by Silvestre et al. (Chap. 11) or in human pathogens as stressed by Rumy et al. (Chap. 10). Illustrations of these new challenges were given on specific diseases or pathogens such as tuberculosis by Lanotte (Chap. 7), pneumocystosis by Chabe et al. (Chap. 8) or hantaviruses by Blasdell et al. (Chap. 9).

In this final conclusion, we will present four general perspectives emerging from chapters.

15.1 The Challenging Development of New Molecular Tools

New typing technologies, such as high-throughput genomics typing, appear and improve at an accelerating rate. This represents a substantial challenge as emphasized by Hallin et al. (Chap. 2), both by the need to deal with a large amount of molecular

S. Morand (✉)
Institut des Sciences de l'Evolution, CNRS-IRD-UM2, CC65, Université Montpellier 2, F-34095 Montpellier, France

Cirad AGIRs, F-34398 Montpellier, France
e-mail: serge.morand@univ-montp2.fr

J. Cabaret
INRA, UR1282 Infectiologie Animale et Santé Publique, F-37380 Nouzilly, France
e-mail: jacques.cabaret@tours.inra.fr

F. Beaudeau
INRA, Oniris, UMR1300 Bio-agression, épidémiologie et analyse de risque en santé animale, BP40706, 44307 Nantes, France
e-mail: francois.beaudeau@oniris-nantes.fr

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2_15,

information and by the integration of these analytical tools in public health laboratory. Tibayrenc (Chap. 3) cited pathogen profiling as one of the new strategies emerging from these new tools, which aims at integrating molecular data (genome, transcriptome, proteome, metabolome) with clinical and epidemiological data, assisted by geographic information systems (GIS).

15.2 Incorporating Concepts of Population and Evolutionary Genetics

Tibayrenc (Chap. 3), Chevignon et al. (Chap. 4), and Charbonnel and Cosson (Chap. 12) stressed the need to incorporate population genetics' concepts in molecular epidemiology.

Chevignon et al. (Chap. 4) emphasized that population genetics may help, among others, characterize the spatial limits of a parasite population and detect drastic changes in parasite population sizes. Hence, population genetics may help understand the population biology of the infectious agents, which may avoid spurious conclusions based on the interpretation of easy-made phylogenetic trees (but see also Archie et al. 2009).

Charbonnel and Cosson (Chap. 4) stressed on some gaps that remained to be assessed to understand the evolution of host resistance gene/polymorphism and adaptation, particularly concerning populations in non-equilibrium such as invasive species. They also referred to the concepts and results of immunoecology, a discipline that aims to explain the variability of immune responses in natural populations, that may fruitfully integrate advances in the molecular epidemiology.

15.3 Incorporating Concepts of Theoretical and Population Modeling

Modeling approach may help for various practical purposes, including evaluation of control strategies as emphasized by Ezanno et al. (Chap. 5). A major challenge is to include genetics in epidemiological modeling. However, as stressed by Ezanno et al. (Chap. 5) classical tools of epidemiological modeling may not be relevant and the development of specific tools is then required to integrate evolutionary ecology and epidemiological patterns.

Similarly, Ferdy and Gandon (Chap. 6) advocated that theoretical models on pathogen virulence evolution will gain considerable predictive power if they can be combined to the approaches developed by molecular biologists that relate virulence, transmission and recovery.

From a different perspective, Caron et al. (Chap. 14) advocated the need of integrating theoretical community ecology such as network-based modeling as tool for predicting emergence avenues. They stressed also the need to integrate molecular information and to develop specific tools for manipulating complex networks.

15.4 From Molecular Epidemiology to Phylogeography of Diseases

The development of new phylogenetic methods has allowed the investigation of pathogens' diversity and their diversification in a spatial context. Phylogeography is a recent science that aims to explain the actual geographic distribution of species and populations by investigating the processes that govern the geographical distributions of lineages within and among closely related species following. Numerous studies have concerned free-living animals and plants, but phylogeographic studies of pathogens and parasites have only recently grew up (Holmes 2008; Nieberding and Olivieri 2007). New methodologies have been recently proposed implementing Bayesian approach for inference, visualization and hypothesis testing of phylogeographic history (Lemey et al. 2009) or in phylodynamics (Holmes and Grenfell 2009). Phylodynamics intend to associate epidemiological dynamics and evolution of pathogens. It has proven an efficient tool for understanding the dynamics of viral infections (Grenfell et al. 2004) but still remains unexplored for bacterial and parasitological infections. Incorporation of molecular data into Geographic Information Systems (GIS) has been made easy following the development of specific softwares.

Molecular epidemiology is a rapidly developing science as it relies on technologies (informatics and high-throughput genomics typing among others). We must take care that conceptual issues are not forgotten and it is one aim of the present book.

References

Archie EA, Luikart G, Ezenwa VO (2009) Infecting epidemiology with genetics: a new frontier in disease ecology. Trends Ecol Evol 24:21–30

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science 303:327–332

Holmes EC (2008) Evolutionary history and phylogeography of human viruses. Annu Rev Microbiol 62:307–328

Holmes EC, Grenfell BT (2009) Discovering the phylodynamics of RNA viruses. PLoS Comput Biol 5(10):e1000505

Lemey P, Rambaut A, Drummond AJ, Suchard MA (2009) Bayesian phylogeography finds its roots. PLoS Comput Biol 5(9):e1000520

Nieberding C, Olivieri I (2007) Parasites: proxies for host genealogy and ecology? Trends Ecol Evol 22:156–165

Index

S. Morand et al. (eds.), *New Frontiers of Molecular Epidemiology of Infectious Diseases*,
DOI 10.1007/978-94-007-2114-2,