

# Infecting epidemiology with genetics: a new frontier in disease ecology

Elizabeth A. Archie<sup>1,2</sup>, Gordon Luikart<sup>1,3</sup> and Vanessa O. Ezenwa<sup>1</sup>

<sup>1</sup> Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA

<sup>2</sup> Current address: Biology Department, Fordham University, 441 E. Fordham Road, Bronx, NY 10458, USA

<sup>3</sup> Centro de Investigação em Biodiversidade e Recursos Genéticos, University of Portugal, Vairão 4485-661, Portugal

**Disease ecologists strive to understand the causes and consequences of parasite infection, including the emergence, spread, persistence and evolution of infectious disease. These processes can be illuminated by parasite genetic markers, which can be used to track parasite spread and infer population history. Recently, a growing number of studies have used molecular tools to examine questions on the ecology of infectious diseases. We review this burgeoning area of research by focusing on three topics where genetic tools will increasingly make major contributions: inferring parasite transmission, reconstructing epidemiological history and identifying physical and environmental drivers of disease spread. We also discuss areas for future research and highlight the promise of multidisciplinary collaborations among geneticists, ecologists and epidemiologists.**

## The value of genetics to disease ecology

Molecular and computational tools from population genetics and phylogenetics hold enormous promise for disease ecology. This is in part because the research problems that concern phylogeneticists and population geneticists—for example, using molecular genetic data to infer population history, understand migration and gene flow and predict evolutionary dynamics—are analogous to processes necessary to understand the ecology and evolution of parasites. Specifically, molecular approaches enhance research in disease ecology because they enable the reconstruction of evolutionary relationships between parasites on a wide range of spatial scales—ranging from within individual hosts to between geographic locations. This fundamental feature improves our ability to track parasite movements, identify parasite origins and understand environmental factors influencing their spread. At a minimum, molecular tools complement and corroborate traditional, epidemiological approaches; at best, these tools greatly improve the resolution of epidemiological methods and allow researchers to address new questions that would be difficult or impossible using traditional epidemiological tools alone.

Although researchers have been applying molecular markers to parasite populations for well over a decade (e.g. [1–4]), in the last few years, new statistical methods [5–9] and increasing collaboration between disease ecologists and population geneticists [10–14] have greatly

## Glossary

**Assignment test:** a statistical test of the hypothesis that the multilocus genotype of an individual arose in a particular population. Sometimes refers to methods that cluster individuals into groups that are genetically related or randomly mating (see Table 2).

**Basic reproductive number ( $R_0$ ):** the average number of secondary infections derived from a single infection in an entirely susceptible population. The value of  $R_0$  determines whether a parasite can invade a host population and how fast it spreads upon invasion.

**Bayesian:** a framework of statistical inference that begins with prior distributions for model parameters and updates these based on observed data to arrive at a posterior probability distribution.

**Coealescent:** a theory that describes the genealogy of chromosomes or genes. Relevant to parasites, this theory describes the shape of a gene genealogy (i.e. the statistical distribution of its branch lengths) under different demographic histories (e.g. exponential growth, stasis or population bottlenecks) in order to date a most recent common ancestor or infer population growth rates (see also Box 1).

**Disease ecology:** the study of interactions between parasites and their hosts, including parasite transmission dynamics; factors underlying patterns of variation in infection; parasite effects on host behavior, population dynamics and community structure; and coevolutionary relationships between hosts and parasites.

**Effective population size:** the size of an ideal population (as defined by population genetics; a stable population with random mating, random variation in reproductive success, equal sex ratio, and nonoverlapping generations) that would experience the same rate of genetic change through genetic drift as the observed population.

**Epidemiological history:** important events and processes occurring in parasite population dynamics, especially the timing of disease outbreaks, and the rate and timing of the parasite population growth rate over the course of one or more epidemics.

**Gene flow:** the movement of genetic material from one population to another, also referred to as migration.

**Landscape ecology:** the study of interactions between spatial patterns and ecological processes, and the spatial extent and configuration of ecological processes. A focus is on understanding the relationship between spatial heterogeneity (anthropogenic-induced and natural) and ecological processes.

**Landscape epidemiology:** assesses the influence of landscape features, environmental variables and spatial heterogeneity on disease spread. Often combines temporal data, spatial data and modeling to predict patterns of disease transmission.

**Landscape genetics:** the use of molecular tools to study how landscape features and environmental variables influence (i) gene flow and the movement of organisms, and (ii) the spatial distribution of genetic diversity at both neutral and functional genetic loci. Landscape genetics is similar to phylogeography (a combination of phylogenetics and biogeography), but is especially useful at finer spatial scales (e.g. when individual hosts or parasites are the unit of study), and for macroparasites, because many of the analytical methods can accommodate multilocus genotypes and nonequilibrium population genetic assumptions.

**Least-cost modeling:** a statistical modeling approach that uses raster-based data inputs to measure the effective distance and connectivity between habitat patches or other geographic areas.

**Migration:** the movement of individuals from one genetically distinct population to another, resulting in gene flow. Migration rate is computed as the probability that a randomly chosen individual (or allele) in each population is a migrant.

**Multilocus genotype:** the combination of the genotypes at each of multiple genetic loci in an individual. Common genetic markers that are used to construct multilocus genotypes include microsatellites and single-nucleotide

Corresponding author: Archie, E.A. (earchie@fordham.edu).

polymorphisms (SNPs).

**Parasite:** disease-causing organisms that live in or on a host, and derive nutrients from this host. Our definition of this term encompasses both microparasites (e.g. viruses, bacteria, protozoa, fungi) and macroparasites (e.g. arthropods, helminths).

**Phylogenetics:** the study of evolutionary relationships among taxa or genetic lineages (e.g. populations or species). Uses analytical techniques to reconstruct the evolutionary tree (phylogeny) with nodes representing taxa or lineages (ancestral or derived), and branch lengths often corresponding to the amount of divergence between groups. Useful for understanding recent transmission processes in rapidly evolving parasites (e.g. RNA viruses); for slowly evolving parasites (e.g. helminths), phylogenetics is useful for inferring long-term evolutionary processes (historical host-switching events) or evolutionary relationships on large geographic scales (e.g. phylogeography).

**Population genetics:** the study of the allele frequency distribution and change influenced by the four evolutionary forces: natural selection, genetic drift, mutation and gene flow. Develops analytical techniques to draw inferences about single populations or metapopulations. Because population genetics can accommodate multilocus genotypes and nonequilibrium assumptions, these techniques are especially useful for understanding evolutionary relationships between macroparasite individuals and populations.

**Transmission:** the process by which susceptible hosts acquire parasites. The rate of transmission depends on the contact rate between hosts, the probability that a contact is with an infectious host and the transmission probability given contact. Our use of transmission here also refers to the movement of parasites between host populations or geographic locations (Table 1).

expanded the breadth of research questions and the range of parasite species that can be studied. In particular, until recently, molecular epidemiology has tended to focus on rapidly evolving RNA viruses; however, two recent advances—the ability to genotype large numbers of genetic loci in the same parasite [15,16] and analytical tools to interpret these multilocus genotypes [7,17,18]—mean that studies at the finest ecological scales are now possible for more slowly evolving parasites, including most macroparasites (see [Glossary](#) under parasite). Furthermore, advances in molecular biology have improved methods for obtaining parasite genetic material; although genetic sampling from parasites can be challenging, particularly from wild hosts, a variety of new techniques allow for extraction and characterization of micro- and macroparasite genetic material from sources ranging from host feces (e.g. [19,20]) to various tissues (e.g. [21]). Here we review three areas of particular interest to disease ecologists: inferring parasite transmission, reconstructing epidemiological history and identifying environmental factors influencing disease spread.

### Inferring parasite transmission

Parasite transmission is arguably the most important process in disease ecology, yet transmission modes can be challenging to uncover and transmission rates are notoriously difficult to quantify [22]. Ecologists and epidemiologists increasingly combine field, experimental and modeling approaches to address several questions related to parasite transmission (Table 1). In addition, phylogenetics and population genetics offer a diverse array of tools to help elucidate transmission processes (Table 2). For any given parasite, the choice of molecular method depends on three main factors: the ecological and evolutionary scale of the research question, how rapidly the parasite evolves and the assumptions of the genetic method. In general, for rapidly evolving parasites (e.g. RNA viruses), phylogenetic tools will be best for inferring transmission, whereas for more slowly evolving parasites (e.g. helminths), transmission will often be best inferred using multilocus

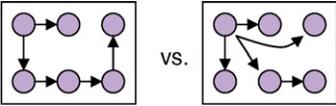
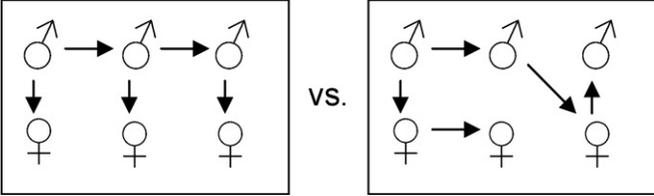
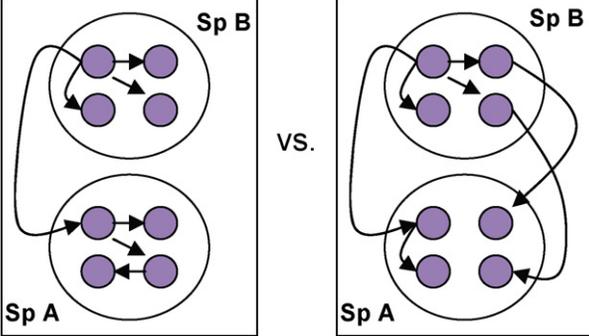
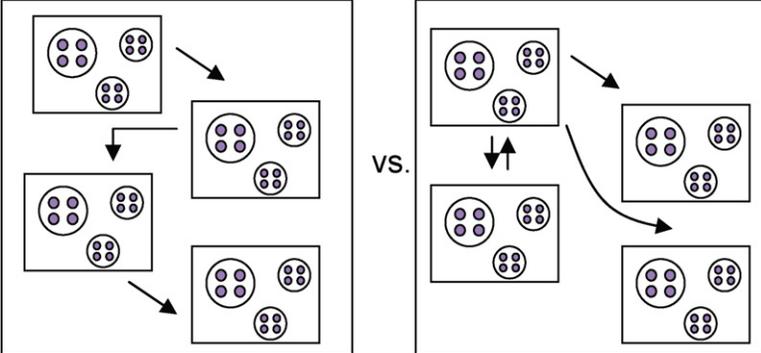
population genetic tools (e.g.  $F_{ST}$ , assignment tests, kinship estimators; see [Table 2](#)).

A single transmission event from one host to another generates a genetic prediction: the genotype(s) of the transmitted parasites must be either a subset of the genotypes contained within the transmitting host or most closely related to the parasites in the transmitting host. This rationale can be expanded to infer the specific geographic pathways over which pathogens are moved by their hosts. For example, global pathways of H5N1 influenza virus (i.e. bird flu) transmission over the last decade were traced by building a phylogeny of H5N1 samples, mapping sample locations onto the tips of the phylogeny and then reconstructing the most parsimonious population of origin for each infection [23]. This phylogenetic approach reconfirmed the major thrusts of H5N1 global diffusion, and for the first time pinpointed the prime source of H5N1 as Guangdong Province, China. On a smaller geographic scale, pathways of foot-and-mouth disease transmission between 20 farms within 100 km<sup>2</sup> were described by using a novel maximum likelihood approach which combined epidemiological and genetic data to resolve pathways of transmission that could not be resolved by either data set alone [24]. Both of these studies demonstrate that statistical phylogeography can identify and test the significance of specific pathways of disease transmission across a variety of geographic scales. Furthermore, the fine-scale dispersal patterns generated by these tools provide valuable information on locations of key disease transmission pathways which can then be targeted by surveillance and intervention programs, greatly enhancing disease control efforts.

Concern about emerging diseases and zoonoses has also stimulated research on parasite transmission between alternative host species and between reservoir and spillover hosts. A variety of genetic tools can address questions of cross-species transmission for both micro- and macroparasites (Table 2). In the simplest cases, phylogenies can identify potential reservoirs of emerging diseases, as in Ebola [25], or pinpoint historical host-switching events for parasites ranging from malaria [26] to human pubic lice [27]. In addition, host and parasite phylogenies can be compared to assess the occurrence of multispecies transmission; for instance, the observation that evolutionary relationships among the nematode parasite *Longistrata caudabullata* were not aligned with the phylogeny of their host species indicated that this nematode is commonly transmitted between two species of voles [28] (see tests of phylogenetic concordance, Table 2). Transmission rates between species can also be inferred quantitatively; for example, assignment tests (Table 2) estimated that 4–7% of *Ascaris* roundworms that infected humans had hybridized with worms infecting pigs [29]. Such studies illustrate the diversity of molecular tools available to elucidate cross-species transmission.

Even on the finest ecological scales, genetic tools can be used to reconstruct chains of transmission and document ‘who infected whom,’ as in a case where phylogenetic methods were used to show that a doctor used a patient’s blood products to infect his former girlfriend with HIV-1 [30]. Within host populations, the distribution of genetic variance within and between populations of parasites

**Table 1. Transmission-related questions commonly addressed by disease ecologists**

Scale of analysis	Questions	Examples	References
Between individual hosts	Who infects whom?  Which individuals contribute most to transmission? Are there heterogeneities in transmission?		[71]
Between groups of individuals (e.g. sex or age classes)	Are there sex biases in transmission?  Are certain age classes more important for disease spread?		[72,73]
Between host populations or species	What is the relative rate of within- versus between-species transmission? Which populations act as sources or sinks of infection? Which populations or species are reservoirs of infection?		[74,75]
Between geographic locations <sup>a</sup>	What are the geographic pathways of transmission? Where do new diseases originate?		[23]

<sup>a</sup>Shaded circles represent host individuals, circles represent host populations and squares represent geographically distinct locations.

infecting individual hosts can reveal heterogeneities in transmission (e.g. AMOVA and  $F_{ST}$ ; see Table 2). For example, the absence of strong genetic differentiation between the parasites infecting different hosts indicated that the transmission dynamics of the helminth *Strongyloides ratti* were dominated by infection by conspecifics rather than by self-reinfection [31]. Furthermore, because highly infected individuals did not reinfect themselves, the observed variability in infection intensities in the host population was likely generated by differences in susceptibility. Despite such key insights derived from this type of work, the use of genetic tools in studies of parasite transmission dynamics within host populations is rare. However, fine-scale genetic data on parasites can be used to identify host individuals that contribute disproportionately to infecting conspecifics; test whether relatives are more likely to infect one another; and link heterogeneities in parasite transmission to individual- and group-level host traits. These questions can be explored for both micro- and macroparasites using tools such as assignment tests, kinship estimators and Bayesian models of coalescent processes [7,17,32,33] (Table 2).

Finally, it is important to note that the studies above all rely on selectively neutral loci (i.e. loci where genetic variants have no difference in darwinian fitness). Neutral loci provide the most accurate representation of transmission patterns because they evolve in relatively predictable ways and should not influence transmission. By contrast, functional or adaptive markers, because they experience natural selection, are likely to obscure transmission patterns or yield biased estimates of transmission rates. However, measuring population genetic structure and gene flow at both neutral and functional loci can lend insight into how parasites adapt to changing environments (e.g. resist drugs, evade host immunity or adapt to climate change). For instance, molecular tools were used to demonstrate that chimpanzees in Kibale National Park, Uganda share *Escherichia coli* bacteria with neighboring human populations—including strains that are resistant to antibiotics [34]. Notably, chimpanzees shared the most bacteria with those humans involved in chimpanzee conservation (e.g. researchers, tourism workers), suggesting that conservation efforts aimed at protecting wildlife can sometimes increase disease risks. Together, the diversity

**Table 2. Major types of phylogenetic and population genetic tools that can be used to infer parasite transmission processes**

Genetic tool, description and common genetic markers	Transmission processes inferred (and examples)	Rationale	Methods <sup>a</sup>
<p><b>Phylogenetics, phylogeography and ancestral character state reconstruction</b> Establishes evolutionary relationships between different taxa. Assumes a bifurcating pattern of evolution. Interprets the phylogeny in geographic space. Reconstructs the character state of ancestral nodes in a phylogeny. Generally applied to molecular (DNA) sequences.</p>	<p><b>Rapidly evolving parasites<sup>b</sup></b></p> <ul style="list-style-type: none"> <li>Identify the presence and direction of transmission between host individuals [30]</li> <li>Infer the most recent population of origin, document recent host-switching events or identify parasite reservoir populations [23,25,75]</li> <li>Draw pathway of transmission among locations [23]</li> <li>Test whether transmission between host populations occurs more often than expected by chance [23,34]</li> <li>Infer the relative rate of transmission between different geographic locations [23]</li> </ul> <p><b>Slowly evolving parasites<sup>c</sup></b></p> <ul style="list-style-type: none"> <li>Document historical host-switching events [26,27]</li> <li>Reconstruct populations of origin and transmission pathways over long timescales</li> </ul>	<p>Parasite genotypes from the recipient host(s) should be nested within the phylogeny of parasite genotypes from the transmitting host(s). Once a phylogeny is built: reconstruct the most likely population of origin for a parasite population; draw pathways of transmission by tracing populations and phylogenetic relationships on a map; test observed phylogenetic relationships to random expectations to understand whether patterns of transmission between host populations are different than expected under random chance; infer relative rates of transmission between locations by counting the number of times parasites moved between ancestral and current population locations.</p>	[76]
<p><b>Phylogenetic network reconstruction</b> Reconstructs evolutionary relationships among haplotypes in an interbreeding population. Usually applied to molecular sequence data, but can also be applied to microsatellites.</p>	<p><b>Rapidly and slowly evolving parasites</b></p> <ul style="list-style-type: none"> <li>Reconstruct genetic relationships between parasite individuals and populations [24]</li> <li>Draw pathways of transmission, potentially between fine-scale geographic locations [24]</li> </ul>	<p>Appropriate for reconstructing evolutionary relationships between parasites (individuals or populations) when haplotypes differ by only one to a few mutational steps. Not restricted to bifurcating branches (allows multiple relationship branch connections; can accommodate recombination).</p>	[77]
<p><b>Tests of phylogenetic concordance</b> Compares congruence between tree topologies (e.g. between host and parasite trees). Usually applied to DNA sequences.</p>	<p><b>Slowly and perhaps rapidly evolving parasites</b></p> <ul style="list-style-type: none"> <li>Test alternative hypotheses about parasite origins [26]</li> <li>Test whether parasites have historically been transmitted between two host species [28]</li> </ul>	<p>If hosts and parasites have coevolved, their phylogenies should be congruent. If parasites are transmitted between host species, then parasite genotypes will not be clustered by host species.</p>	[76,78]
<p><b>F-statistics; Analysis of Molecular Variance (AMOVA)</b> Measures the proportion of genetic variation that occurs within and between different populations by using multilocus allele frequency and or allele genealogy data. Can be applied to almost any genetic marker, including DNA sequences, fragment length polymorphisms, microsatellites or SNP genotypes.</p>	<p><b>Slowly and rapidly evolving parasites</b></p> <ul style="list-style-type: none"> <li>Infer relative rates of transmission within and between parasite subpopulations (e.g. individual hosts, host demographic groups, social groups, host species) [31,34]</li> <li>Infer pathways of transmission between parasite populations</li> </ul>	<p>This equilibrium population genetic approach measures the combined effects of migration and genetic drift over time. Differences in <i>F<sub>ST</sub></i> can be interpreted as differences in average rates of migration between populations over the past 10s to 1000s of generations depending on <i>N<sub>e</sub></i> and <i>N<sub>m</sub></i> (i.e. the effective migration rate). Draw pathways of transmission by tracing populations and genetic distances on a map.</p>	[7]
<p><b>Assignment tests</b> A statistical test of the hypothesis that the multilocus genotype of an individual arose in a particular population. Usually applied to multilocus microsatellite genotypes.</p>	<p><b>Slowly evolving parasites</b></p> <ul style="list-style-type: none"> <li>Assign individual parasites to a most likely population of origin (e.g. population with highest expected frequency of a multilocus genotype) [29,79]</li> <li>Infer transmission rates between populations [29]</li> </ul>	<p>This nonequilibrium population genetic approach directly documents recent migration events. Infer transmission rates by counting or inferring the number of individuals in one subpopulation that were derived from a different subpopulation.</p>	[7,33]

Table 2 (Continued)

Genetic tool, description and common genetic markers	Transmission processes inferred (and examples)	Rationale	Methods
<p><b>Landscape genetic analyses</b></p> <p>Identifies landscape features and environmental variables correlated with genetic distance between individuals or populations. Can be applied to almost any genetic marker, especially molecular sequences or microsatellites.</p>	<p><b>Rapidly and slowly evolving parasites</b></p> <ul style="list-style-type: none"> <li>• Test whether certain features (roads, rivers, forests, open areas, etc.) are correlated with genetic distance, genetic relatedness or gene flow [48]</li> <li>• Test among 10s to 1000s of alternative hypotheses (i.e. resistance maps) to identify the combination of landscape resistance values that best explain (correlate with) gene flow or genetic distance</li> </ul>	<p>This nonequilibrium approach identifies locations on the landscape where genetic discontinuities occur or gene flow is low and which represent 'barriers' between populations. It also identifies locations of high gene flow representing corridors of extensive movement or high transmission.</p>	[8,18,47]

<sup>a</sup>See also <http://evolution.genetics.washington.edu/phyloip/software.html> and <http://www.biology.lsu.edu/general/software.html> for software.

<sup>b</sup>Most microparasites (e.g. RNA viruses, DNA viruses, bacteria) tend to be rapidly evolving, but it is important to remember that some microparasites might undergo slowly evolving phases (e.g. bacterial spores that remain dormant for long periods of time).

<sup>c</sup>Slowly evolving parasites tend to include most macroparasites (e.g. helminths, arthropods).

of molecular tools outlined above has the potential to make major contributions to understanding transmission processes.

### Reconstructing epidemiological history

Genetic tools offer new approaches to infer epidemiological history, especially to date parasite introductions, infer changes in parasite population size over time and estimate the parasite basic reproductive number ( $R_0$ ). Most of these methods are based on coalescent theory and use Bayesian models to resolve a large number of complex evolutionary processes (Box 1). These tools can be applied to questions on various timescales—from weeks to millennia—and a range of parasite taxa, including micro- and macroparasites; however, analytical assumptions will vary depending on the parasite's evolutionary rate, effective population size and generation time. Typically, Bayesian coalescent models have been applied to viruses which, owing to their high mutation rates, large population sizes and periodic

#### Box 1. Taming the BEAST: using Bayesian statistics and coalescent theory to estimate $R_0$

Coalescent theory is based on the observation that all the alleles in a given population can be traced, through the coalescence of lineages in a genealogy, back to a single, most recent common ancestor. Both the rate of coalescence and its pattern are influenced by population size (Figure 1). For instance, where the size of the population does not change and individuals reproduce at random (Figure 1a), lineages coalesce at a constant rate. However, when population sizes are small (or declining), lineages coalesce more quickly (Figure 1b), and when population sizes are large (or expanding), coalescence slows (Figure 1c) (see Ref. [58]).

Bayesian coalescent models capitalize on changes in the rate and pattern of coalescence to infer population demographic dynamics. These models (reviewed in Ref. [6]) can accommodate a variety of genetic markers and sampling schemes, from microsatellites to molecular sequences. In general, coalescent models estimate demographic parameters by: (i) constructing a genealogical tree from a set of genetic marker data; (ii) generating a set of expected trees from the data under the assumptions of various evolutionary models; and (iii) testing the observed tree against the expected trees to find the evolutionary model, and its associated demographic parameters, that best fit the data.

For parasites, one commonly used Bayesian coalescent model is included in a software package called BEAST [5]. BEAST is unique in that it takes advantage of genetic samples collected at multiple points in time to infer how those sequences evolved, and in turn infer demographic parameters such as changes in effective population size (i.e. the effective number of infections) and population divergence times. Of particular value to epidemiologists, BEAST can be used to estimate  $r$ , the parasite population growth rate in a wholly susceptible population. BEAST calculates  $r$  from inferred changes in the parasite effective population size ( $N_e$ ) over time. This rate of change,  $r$ , can be converted to  $R_0$ , the basic reproductive number, using the following equation [44]:

$$R_0 = rD + 1,$$

where  $D$  is the average duration of parasite infectiousness [41,44].

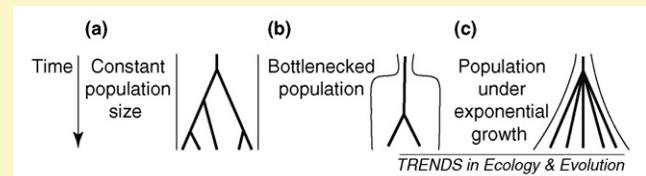


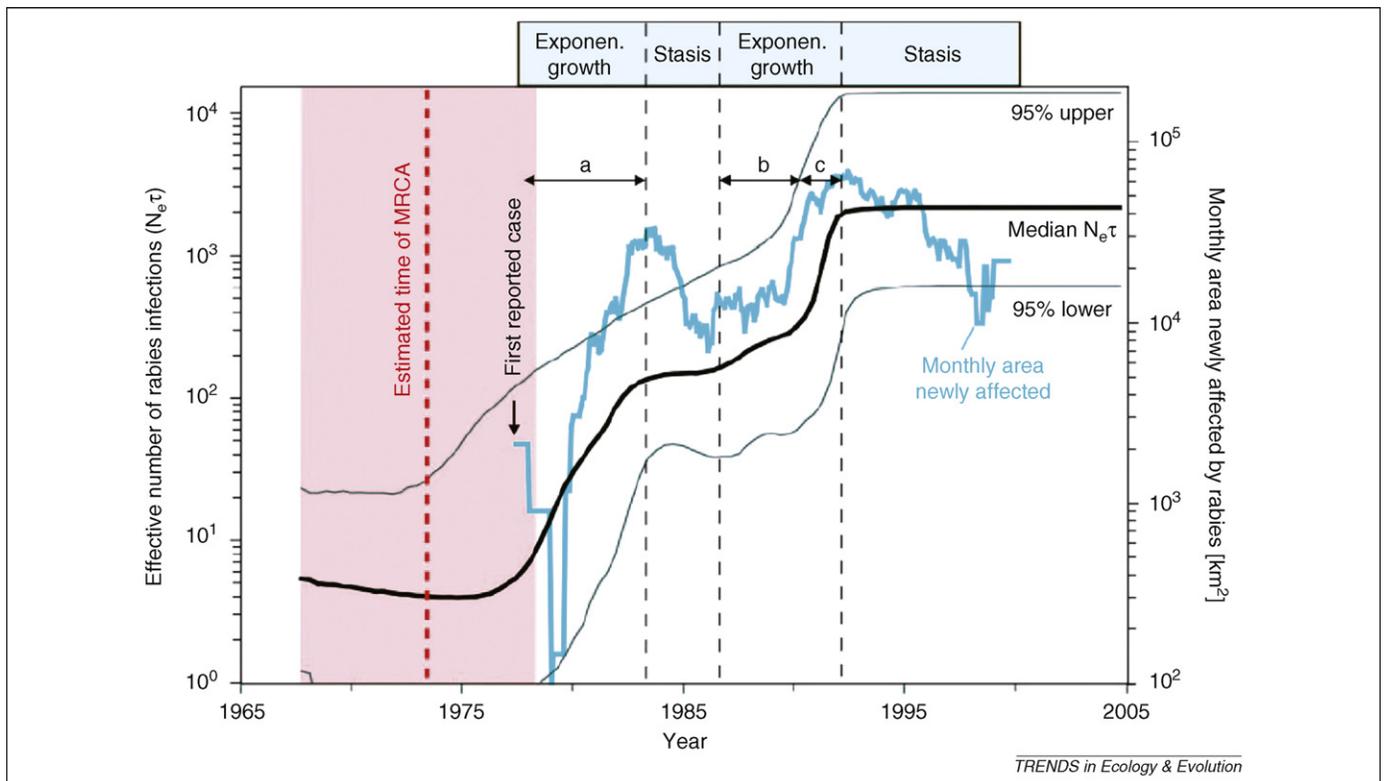
Figure 1. Changes in population size affect the rate of coalescence. Branches indicate parasite genotypes.

transmission bottlenecks, experience ecological and evolutionary changes on comparable timescales. Fewer studies have applied coalescent models to macroparasite populations (but see Ref. [35]); however, such studies are feasible with multilocus statistical frameworks [7,36].

One use of Bayesian coalescent models that has brought new insights to disease ecology is the ability to date parasite introductions or disease outbreaks that occurred hundreds or even thousands of years ago [35,37–39]. Such information has been used to test hypotheses about factors that contribute to disease introductions and emergence—for instance, to understand anthropogenic causes of disease spread. Coalescent models were used to understand the role of human migrations on the origins of yellow fever virus (YFV) in the Americas [38]. YFV is thought to have originated in Africa and might have been brought to the Americas with its mosquito vector in the bilges of slave-trading vessels. Consistent with this hypothesis, YFV was nested within the African YFV phylogeny, and African and American YFV last shared a common ancestor ~300–400 years ago. Notably, this deep divergence indicates that vaccination has been effective in preventing more recent movement of YFV between Africa and the Americas, despite continued human migration. Similarly, coalescent models were used to test whether the colonial period in Africa contributed to the emergence of rift valley fever

(RVF) [39]. In support, the time to most recent common ancestor (TMRCA) for RVF was dated to the 1800 s, a time of dramatic changes in traditional agriculture and the introduction of nonnative cattle breeds.

TMRCA estimates, however, should be interpreted with caution because they will often date only the most recent population bottleneck. For instance, TMRCA estimates for measles, mumps and canine distemper are all within the last hundred years, even though these viruses have probably been infecting host populations for centuries [40]. These TMRCA estimates were likely more recent than expected because the effective population sizes of these viruses decrease dramatically between epidemics, increasing the likelihood of coalescence. Hence, for parasites that undergo large fluctuations in population size, coalescent methods are likely to date only the most recent population expansion. Similarly, if TMRCA estimates are estimated using loci under selection but the coalescent model assumes neutrality, then the resulting estimate is likely to be inaccurate. For instance, because neutral mutations accumulate over time, a locus that has experienced diversifying selection might appear much older than it actually is, whereas a locus under purifying selection might appear more recent. However, some coalescent models accommodate selection, and these models might be useful for understanding when and how selection has acted on genes of functional interest.



**Figure 1.** The epidemiological history for raccoon rabies virus (RRV) in the northeastern United States, depicted as the change in the number of RRV infections over time (1977–2005). The graph shows RRV epidemiological history estimated from both genetic and epidemiological data. The blue line represents the epidemiological index of RRV population history; this line is the moving average (for a 15 month span) of the geographic area (in km<sup>2</sup>) that was newly affected by rabies between 1977 and 1999. The thick black line is the genetic estimate of RRV population history; it represents the median effective number of rabies virus infections calculated using Bayesian coalescent modeling (BEAST software [5]). Specifically, this line is calculated as the product of the virus effective population size ( $N_e$ ) and generation time ( $\tau$ ) in years. The thin black lines represent 95% highest posterior density intervals. The red dashed line is the Bayesian estimate for the date of the most recent common ancestor (MRCA); the shaded pink area is the 95% highest posterior density intervals for the MRCA estimate (figure is reproduced with permission from Ref. [41]). The close congruence between genetic and epidemiological estimates of RRV population history validates the accuracy of molecular approaches for understanding epidemiological history. The MRCA estimate was consistent with the date of the first documented RRV cases in 1977. In addition, the genetic and epidemiological estimates for the mode of population growth were nearly identical; both genetic and epidemiological modes of inference indicate that the number of infections did not increase at a uniform rate, but rather occurred during periods of exponential growth, followed by stasis (i.e. stable population size).

In addition to dating parasite outbreaks, Bayesian coalescent models can also help researchers identify the mode and rate of parasite population growth. For example, molecular sequences from serially collected virus samples were used to reconstruct the epidemiological history (from 1977 to 2005) of the raccoon rabies virus in the north-eastern United States (Figure 1 [41]). This study showed that the number of rabies infections did not increase at a uniform rate, but instead occurred during periods of exponential growth, followed by stable population size. Similar genetic analyses have confirmed that the demography of emerging parasites is often best modeled by logistic growth, and they provide a more detailed picture of evolutionary and ecological processes that occur during emergence [37,42,43]. Specifically, these analyses show that during emergence, the population genetic structure of many parasites is characterized by an invasion phase, where parasite genetic diversity increases rapidly, and a maintenance phase, where genetic diversity remains fairly constant [37,39,43].

Because coalescent approaches can reconstruct patterns of parasite population growth, these tools have been used to compare growth rates in different strains of the same virus [37,43]. Such data are useful for testing hypotheses about competition between parasite strains. For instance, the emergence of canine parvovirus (CPV) was characterized by unusually rapid adaptive evolution of a new virus strain and between-strain competition [42]. CPV emerged from feline panleukopenia virus and early strains lacked the ability to infect felids. However, the latest strain, CPV2a, is rapidly replacing older strains and, unlike them, CPV2a has evolved a broader host range and is able to infect felids [42]. Between-strain competition appears to be a feature of many parasite populations, and genetic tools will increasingly play a role in elucidating the mechanisms underlying competitive interactions between parasites.

Finally, parasite population growth rates derived from Bayesian coalescent models can be used to estimate  $R_0$ , a key parameter of epidemiological models that determines whether a parasite can invade and establish within a host population and how fast it spreads (Box 1). This method, used to measure the expansion of hepatitis C virus (HCV) in Egypt [44,45], demonstrated that unsterile injection equipment used during anti-schistosomiasis treatment from the 1920 s to 1980 s increased  $R_0$  from 1.7 to between 3 and 7. Similarly, coalescent models were used to calculate  $R_0$  for rabies in the northeastern United States, revealing surprisingly low estimates of  $R_0$  (1.02–1.16), as compared to other host populations [41]. Studies like these and those described above demonstrate the power of Bayesian coalescent models for understanding the history of epidemics and factors influencing parasite fitness.

### Identifying environmental and landscape drivers of disease spread

A major goal of landscape epidemiology is to understand how environmental variables affect the dynamics of infectious disease (reviewed in Ref. [46]). Landscape epidemiology can be advanced by integrating tools from landscape genetics (reviewed in Refs [8,47]), which is an emerging discipline that combines population genetics, landscape

ecology and spatial statistics to assess how landscape features and environmental variables influence individual movement and microevolutionary processes (gene flow, genetic drift and local adaptation). The value of landscape genetics to disease ecology is twofold; first, landscape genetics offers an efficient way to gain an unusually fine scale understanding of where and why parasites are moving on the landscape. Second, landscape genetics offers new insights into parasite evolution by employing both neutral and functional genetic loci to understand the spatial aspects of evolutionary change (e.g. to observe local adaptation or evolution along the advancing edge of an epidemic). However, despite this potential, landscape genetic approaches have seldom been used on parasite populations, although analytical tools are available to conduct such studies for both macro- and microparasites.

Landscape genetics can identify landscape features influencing parasite spread by tracking host and/or parasite movement across multiple spatial scales (Box 2). For instance, landscape genetics was used to identify environmental features influencing the distribution of chronic wasting disease (CWD) in white-tailed deer in Wisconsin,

#### Box 2. Landscape genetics: a promising approach to understanding disease spread

Physical barriers that impede host movement can also reduce the spread of parasites. From its first detection on the Virginia–West Virginia border in the late 1970 s [59], raccoon rabies virus spread rapidly across the northeastern United States, eventually arriving in Canada in 1999. Multiple landscape features limit the spread of raccoon rabies [41,60], but quantifying these effects can be challenging. However, a recent study that assessed the risk of raccoon rabies expansion illustrates how these challenges can be mitigated by a landscape genetic approach [61].

The study used landscape genetics and computer simulations to predict the movement of raccoons and, by extension, rabies, from western New York, across the Niagara River, and into Canada [61]. The authors used an individual-based, spatially explicit model to simulate different success rates of raccoon attempts to cross the Niagara River (i.e. permeabilities). The model also included information about the mitochondrial genotype of each raccoon, and a measure of genetic differentiation ( $\phi_{ST}$ ; similar to  $F_{ST}$  in Table 2) was used to calculate genetic differentiation across the river, under different permeability simulations. To assess the impact of the river, the simulated population genetic structures were compared to the actual population genetic structure, based on mitochondrial DNA sequences from 166 raccoons from across the study range. The best match between simulated and actual genetic structures indicated that the Niagara River prevents 50% of raccoon crossings.

This example highlights how measures of gene flow can calibrate the effect of barriers to host movement and help estimate the impact of the physical environment on the likelihood of disease spread. For raccoon rabies, genetic data from the virus (as opposed to the host) could further reveal whether particular viral genotypes are more likely to immigrate (be transmitted) and whether landscape features exert selective pressure on parasites.

More broadly, the availability of molecular markers for an increasing number of hosts and parasites means that similar studies can now be done in many disease systems. For instance, markers are available for house finch conjunctivitis [62] and rabbit myxoma virus [63], and the spread of both of these parasites is known to be influenced by landscape features (e.g. bird feeders [64] and rivers [65]). The combination of host and parasite genetic data in a landscape genetics framework promises to lend new insight into how landscape features shape the movements of some of the smallest organisms on earth.

USA [48]. Specifically, genetic differentiation between deer populations was greatest, and CWD prevalence lowest, in areas separated by a river, indicating that this landscape feature reduced disease spread. This result suggests that landscape genetics can help predict populations at high risk of infection based on their genetic connectivity to infected host populations and might also help target areas for disease surveillance and preventative measures (e.g. along rivers).

Because landscape genetics can reconstruct parasite movements, it is extremely useful for verifying and improving the resolution of more traditional epidemiological approaches [49–51]. For example, parasite sequence data verified epidemiological estimates of the geographic origin, routes and time period of the initial raccoon rabies outbreak in the northeastern United States (Figure 1 [41]). Mountain ridges and waterways impeded rabies spread by reducing host movement. In addition, genetic data verified the direction of rabies spread from the epizootic origin, and for the first time revealed that each viral genetic lineage was associated with a specific path of spread. These paths were evident 30 years after the initial virus expansion. This study demonstrates the power of landscape genetics to retrospectively (i.e. decades later) identify transmission barriers as well as the origin, timing and direction of parasite spread. The future application of landscape genetics to epidemiological questions will likely be enhanced by new statistical approaches such as least-cost modeling (e.g. [18]), which allows for testing the relative contributions of multiple environmental and landscape variables influencing parasite movement. Such techniques will vastly improve our understanding of how diseases spread across complex landscapes resulting from increased habitat fragmentation and urbanization.

In addition to documenting fine-scale patterns of parasite movement, landscape genetics can be applied to questions of parasite local adaptation and spatial evolutionary change. For instance, genetic data from evolutionarily neutral loci were combined with spatial information to measure evolution in rabies during an epidemic in foxes across southern Ontario, Canada [50]. Divergence among viral lineages correlated with their position along the advancing epidemic; newly derived lineages were along the wave front, and divergent ancestral lineages were near the outbreak origin. Furthermore, the regional population was partitioned into two genetically distinct subpopulations, each corresponding to one of the two arms of the advancing wave of spread of rabies in foxes. The study illustrates that integrating geographic analysis of genetic variants with information on spatial dynamics can reveal patterns of parasite evolutionary change during disease spread.

Finally, questions on parasite local adaptation can be addressed using techniques that assess population genetic structure at both neutral and adaptive gene markers. This multilocus approach—also called population genomics—involves the simultaneous genotyping of multiple loci from the genomes of many parasite individuals across environmental and landscape gradients [16,52,53]. This approach could greatly advance knowledge of adaptive differentiation in parasites, but is seldom used. However, one

study of the bacterium *Campylobacter jejuni* used multiple loci and spatial genetic modeling to describe transmission pathways at fine spatial scales and found evidence for local adaptation to certain niches or habitats, including host species and environmental water sources [54]. This study highlights the utility of landscape genomics for understanding parasite adaptation to environmental variables likely to influence disease spread and persistence.

### Concluding remarks and future directions

Molecular tools have greatly advanced scientific understanding of parasite ecology and evolution. Collaborations between disease ecologists and geneticists are increasingly common [41,48], and the burgeoning use of genetic techniques in disease ecology creates several exciting avenues for future research (Box 3). Although these examples highlight the promise of molecular tools, one current limitation

#### Box 3. Future directions

##### Unraveling heterogeneities in transmission

Genetic techniques are increasingly used to identify sources of infection, whether from individual hosts [30] or locations across the globe [23]. At the same time, disease ecologists are concerned with understanding how variability in infectiousness, across individuals or groups, characterizes infectious disease systems [66]. Identifying superspreaders (i.e. individuals who contribute disproportionately to parasite transmission) is key for effective disease management and control [66], and genetic tools offer a novel way of detecting transmission heterogeneities in both human and wildlife systems, including detecting sex, age or other biases in infectiousness and quantifying differences in the magnitude of these heterogeneities across systems.

##### Understanding the origin and emergence of human diseases

Repeated transmission of parasites between humans and animals is thought to be an important mechanism driving the emergence of zoonotic diseases [67]. Very little is known about the animal origins of diseases that have plagued humans in the past [68], and even less about which parasite species might pose the greatest threats for the future. Genetic techniques can play a central role in clarifying the origins of human diseases and documenting the frequency of cross-species transmission events between humans and animals [29]. For example, do close evolutionary relationships between humans and animals predict animal sources of human diseases? To what degree do species with intimate human associations (e.g. pets, livestock, peridomestic species) contribute to zoonotic infections? What ecological and anthropogenic factors are associated with frequent parasite exchange between humans and animals? What is the genetic basis of host specificity?

##### Linking climate, environment and parasite evolution

Climate change, urbanization and habitat fragmentation are likely to be key factors influencing the dynamics of many infectious diseases in the future [69,70]. Whereas disease ecologists commonly address environment–disease interactions by tracking host susceptibility, exposure and infection status, genetic approaches offer an additional dimension—tracking evolutionary changes in parasites. Because climate warming or land-use changes can affect key parasite traits such as development rate, transmission efficiency and geographic distribution, these will likely be potent selective forces on parasites. Genetic tools can be used to link information on genetic variation and demographic change in parasite populations (e.g. Box 1) with information on climate or habitat and host infection rates [53]. Such approaches can be used not only to understand parasite evolutionary change in response to changing climate, land use or other environmental conditions in the future but also to detect changes that might have already occurred in the past using ancient parasite DNA [55].

is the lack of suitable molecular markers for many parasite systems. Thus, there is a critical need for widespread development of neutral and candidate gene markers for more parasite species. Encouragingly, the recent explosion of new molecular techniques, including massively parallel sequencing (e.g. pyrosequencing [15,55]), multiplex PCR amplification of thousands of loci simultaneously [56], quantitative PCR and microarrays [57] and ongoing parasite genome sequencing projects (e.g. see <http://www.sanger.ac.uk/Projects/Pathogens>), foreshadows a revolution in this area. For example, massively parallel sequencing (e.g. metagenomics) promises extremely detailed information on parasite communities, rich data sets on patterns of evolutionary change and abundant genetic markers [21]. Disease ecologists and geneticists are well positioned to galvanize this movement and forge, together, a new frontier in infectious disease ecology.

### Acknowledgements

We thank S. Altizer, T. Cosart, M. Kardos, C. Williams and three anonymous reviewers for valuable comments on the manuscript. G.L. was supported by the Portuguese-American Foundation for Development, CIBIO and UP.

### References

- Blouin, M.S. *et al.* (1995) Host movement and the genetic structure of populations of parasitic nematodes. *Genetics* 141, 1007–1014
- Tibayrenc, M. (1995) Population genetics of parasitic protozoa and other micro-organisms. *Adv. Parasitol.* 36, 48–115
- Nadler, S.A. (1995) Microevolution and the genetic structure of parasite populations. *J. Parasitol.* 81, 395–403
- Amann, R.I. *et al.* (1995) Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* 59, 143–169
- Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214
- Beaumont, M.A. and Rannala, B. (2004) The Bayesian revolution in genetics. *Nat. Rev. Genet.* 5, 251–261
- Excoffier, L. and Heckel, G. (2006) Computer programs for population genetics data analysis: a survival guide. *Nat. Rev. Genet.* 7, 745–758
- Manel, S. *et al.* (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends Ecol. Evol.* 18, 189–197
- Beerli, P. (2006) Genetics and population analysis comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22, 341–345
- Criscione, C.D. *et al.* (2005) Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Mol. Ecol.* 14, 2247–2257
- de Meeus, T. *et al.* (2007) Population genetics and molecular epidemiology or how to “debusquer la bete”. *Infect. Genet. Evol.* 7, 308–332
- Perez-Losada, M. *et al.* (2007) New methods for inferring population dynamics from microbial sequences. *Infect. Genet. Evol.* 7, 24–43
- Paterson, S. and Viney, M.E. (2000) The interface between epidemiology and population genetics. *Parasitol. Today* 16, 528–532
- Criscione, C.D. and Blouin, M.S. (2005) Effective sizes of macroparasite populations: a conceptual model. *Trends Ecol. Evol.* 21, 212–217
- Ellegren, H. (2008) Sequencing goes 454 and takes large-scale genomics into the wild. *Mol. Ecol.* 17, 1629–1631
- Luikart, G. *et al.* (2003) The power and promise of population genomics: from genotyping to genome typing. *Nat. Rev. Genet.* 4, 981–994
- Faubet, P. *et al.* (2007) Evaluating the performance of a multilocus method for the estimation of migration rates. *Mol. Ecol.* 16, 1149–1166
- Cushman, S.A. *et al.* (2006) Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *Am. Nat.* 168, 486–499
- Leendertz, F.H. *et al.* (2006) Pathogens as drivers of population declines: the importance of systematic monitoring in great apes and other threatened mammals. *Biol. Conserv.* 131, 325–337
- Whittier, C.A. *et al.* (2004) Comparison of storage methods for reverse-transcriptase PCR amplification of rotavirus RNA from gorilla (*Gorilla g. gorilla*) fecal samples. *J. Virol. Methods* 116, 11–17
- Cox-Foster, D.L. *et al.* (2007) A metagenomic survey of microbes in honey bee colony collapse disorder. *Science* 318, 283–287
- McCallum, H. *et al.* (2001) How should pathogen transmission be modeled? *Trends Ecol. Evol.* 16, 295–300
- Wallace, R.G. *et al.* (2007) A statistical phylogeography of influenza A H5N1. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4473–4478
- Cottam, E.M. *et al.* (2008) Integrating genetic and epidemiological data to determine transmission pathways of foot-and-mouth disease virus. *Proc. Biol. Sci.* 275, 887–895
- Leroy, E.M. *et al.* (2005) Fruit bats as reservoirs of Ebola virus. *Nature* 438, 575–576
- Mu, J. *et al.* (2005) Host switch leads to emergence of *Plasmodium vivax* malaria in humans. *Mol. Biol. Evol.* 22, 1686–1693
- Reed, D.L. *et al.* (2007) Pair of lice lost or parasites regained: the evolutionary history of anthropoid primate lice. *BMC Biol.* 5, 7
- Brant, S.V. and Orti, G. (2003) Evidence for gene flow in parasitic nematodes between two host species of shrews. *Mol. Ecol.* 12, 2853–2859
- Criscione, C.D. *et al.* (2007) Disentangling hybridization and host colonization in parasitic roundworms of humans and pigs. *Proc. Biol. Sci.* 274, 2669–2677
- Metzker, M.L. *et al.* (2002) Molecular evidence for HIV-1 transmission in a criminal case. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14292–14297
- Paterson, S. *et al.* (2000) Inferring infection process of a parasitic nematode using population genetics. *Parasitology* 120, 185–194
- Waples, R.S. and Gaggiotti, O. (2006) What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Mol. Ecol.* 15, 1419–1439
- Manel, S. *et al.* (2005) Assignment methods: matching biological questions with appropriate techniques. *Trends Ecol. Evol.* 20, 136–142
- Goldberg, T.L. *et al.* (2007) Patterns of gastrointestinal bacterial exchange between chimpanzees and humans involved in research and tourism in western Uganda. *Biol. Conserv.* 135, 511–517
- Cornuet, J.M. *et al.* (2006) Inference on microsatellite mutation processes in the invasive mite, *Varroa destructor*, using reversible jump Markov chain Monte Carlo. *Theor. Popul. Biol.* 69, 129–144
- Beaumont, M.A. (1999) Detecting population expansion and decline using microsatellites. *Genetics* 153, 2013–2029
- Carrington, C.V.F. *et al.* (2005) Invasion and maintenance of dengue virus type 2 and type 4 in the Americas. *J. Virol.* 79, 14680–14687
- Bryant, J.E. *et al.* (2007) Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas. *PLoS Pathog.* 3, e75
- Bird, B.H. *et al.* (2007) Complete genome analysis of 33 ecologically and biologically diverse rift valley fever virus strains reveals widespread virus movement and low genetic diversity due to recent common ancestry. *J. Virol.* 81, 2805–2816
- Pomeroy, L.W. *et al.* (2008) The evolutionary and epidemiological dynamics of the paramyxoviridae. *J. Mol. Evol.* 66, 98–106
- Biek, R. *et al.* (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7993–7998
- Shackelton, L.A. *et al.* (2005) High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc. Natl. Acad. Sci. U. S. A.* 102, 379–384
- Snapinn, K.W. *et al.* (2007) Declining growth rate of West Nile virus in North America. *J. Virol.* 81, 2531–2534
- Pybus, O.G. *et al.* (2001) The epidemic behaviour of hepatitis C virus. *Science* 292, 2323–2325
- Pybus, O.G. *et al.* (2003) The epidemiology and iatrogenic transmission of hepatitis C virus in Egypt: a Bayesian coalescent approach. *Mol. Biol. Evol.* 20, 381–387
- Ostfeld, R.S. *et al.* (2005) Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends Ecol. Evol.* 20, 328–336
- Holderegger, R. and Wagner, H.H. (2008) *Landscape Genetics*. Bioscience
- Blanchong, J.A. *et al.* (2008) Landscape genetics and the spatial distribution of chronic wasting disease. *Biol. Lett.* 4, 130–133
- Real, L.A. and Biek, R. (2007) Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. *J. R. Soc. Interface* 4, 935–948

- 50 Real, L.A. *et al.* (2005) Unifying spatial population dynamics and molecular evolution of epidemic rabies virus. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12107–12111
- 51 Wood, M.J. *et al.* (2007) Within-population variation in prevalence and lineage distribution of avian malaria in blue tits. *Cyanistes caeruleus*. *Mol. Ecol.* 16, 3263–3273
- 52 Holderegger, R. *et al.* (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecol.* 21, 797–807
- 53 Hoffmann, A.A. and Willi, Y. (2008) Detecting genetic responses to environmental change. *Nat. Rev. Genet.* 9, 421–432
- 54 French, N. *et al.* (2005) Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environ. Microbiol.* 7, 1116–1126
- 55 Millar, C.D. *et al.* (2008) New developments in ancient genomics. *Trends Ecol. Evol.* 23, 386–393
- 56 Porreca, G.J. *et al.* (2007) Multiplex amplification of large sets of human exons. *Nat. Methods* 4, 931–936
- 57 Monis, P.T. and Giglio, S. (2006) Nucleic acid amplification-based techniques for pathogen detection and identification. *Infect. Genet. Evol.* 6, 2–12
- 58 Biek, R. *et al.* (2006) A virus reveals population structure and recent demographic history of its carnivore host. *Science* 311, 538–541
- 59 Winkler, W. and Jenkins, S. (1991) Raccoon rabies. In *The Natural History of Rabies* (Baer, G.M., ed.), pp. 240–325, CRC Press
- 60 Smith, D.L. *et al.* (2002) Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3668–3672
- 61 Rees, E.E. *et al.* (2008) Assessing a landscape barrier using genetic simulation modeling: implications for raccoon rabies management. *Prev. Vet. Med.* 86, 107–123
- 62 Ferguson, N.M. *et al.* (2005) Use of molecular diversity of *Mycoplasma gallisepticum* by gene-targeted sequencing (GTS) and random amplified polymorphic DNA (RAPD) analysis for epidemiological studies. *Microbiology* 151, 1883–1893
- 63 Merchant, J.C. *et al.* (2003) Monitoring the spread of myxoma virus in rabbit *Oryctolagus cuniculus* populations on the southern tablelands of New South Wales, Australia. III. Release, persistence and rate of spread of an identifiable strain of myxoma virus. *Epidemiol. Infect.* 130, 135–147
- 64 Dhondt, A.A. *et al.* (2005) Dynamics of a novel pathogen in an avian host: mycoplasmal conjunctivitis in house finches. *Acta Trop.* 94, 77–93
- 65 Ratcliffe, F.N. *et al.* (1952) Myxomatosis in Australia: a step towards the biological control of the rabbit. *Nature* 170, 7–12
- 66 Lloyd-Smith, J.O. *et al.* (2005) Superspreading and the effect of individual variation on disease emergence. *Nature* 438, 355–359
- 67 Wolfe, N.D. *et al.* (2005) Bushmeat hunting, deforestation, and prediction of zoonotic disease emergence. *Emerg. Infect. Dis.* 11, 1822–1827
- 68 Wolfe, N.D. *et al.* (2007) Origins of major human infectious diseases. *Nature* 447, 279–283
- 69 Harvell, C.D. *et al.* (2002) Climate warming and disease risks for terrestrial and marine biota. *Science* 296, 2158–2162
- 70 LeBarbençon, C. *et al.* (2008) Evolution of pathogens in a man-made world. *Mol. Ecol.* 17, 475–484
- 71 Kiss, I.Z. *et al.* (2005) Disease contact tracing in random and clustered networks. *Proc. Biol. Sci.* 272, 1407–1414
- 72 Ferrari, N. *et al.* (2007) The role of sex in parasite dynamics: model simulations on transmission of *Heligmosomoides polygyrus* in populations of yellow-necked mice, *Apodemus flavicollis*. *Int. J. Parasitol.* 37, 341–349
- 73 Perkins, S.E. *et al.* (2003) Empirical evidence for key hosts in persistence of tick-borne disease. *Int. J. Parasitol.* 33, 909–917
- 74 Begon, M. *et al.* (1999) Transmission dynamics of a zoonotic pathogen within and between wildlife host species. *Proc. Biol. Sci.* 266, 1939–1945
- 75 Lembo, T. *et al.* (2008) Exploring reservoir dynamics: a case study of rabies in the Serengeti ecosystem. *J. Appl. Ecol.* 45, 1246–1257
- 76 Felsenstein, J. (2004) *Inferring Phylogenies*. Sinauer Associates
- 77 Posada, D. and Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends Ecol. Evol.* 16, 37–45
- 78 Goldman, N. *et al.* (2000) Likelihood-based tests of topologies in phylogenetics. *Syst. Biol.* 49, 652–670
- 79 Criscione, C.D. *et al.* (2006) Parasite genotypes identify source populations of migratory fish more accurately than fish genotypes. *Ecology* 87, 823–828