Theory of Host–Parasite Coevolution: From Ecology to Genomics

Aurélien Tellier, *Professorship for Population Genetics, Department of Life Science Systems, Technical University of Munich, Freising, Germany*

James KM Brown, *John Innes Centre, Norwich, UK*

Mike Boots, *Department of Integrative Biology, University of California, Berkeley, California, USA*

Sona John, *Professorship for Population Genetics, Department of Life Science Systems, Technical University of Munich, Freising, Germany*

Advanced article

Article Contents

Published online: 26th May 2021

Antagonistic coevolution between hosts and parasites is a ubiquitous process in nature. The theory of coevolution describes how reciprocal changes in genotype frequencies occur in the interacting species over time. Host–parasite coevolution is driven by two types of frequency-dependent selection (FDS): indirect FDS arises from the host–parasite interaction, while direct FDS arises from ecological features which affect coevolution. Depending on the balance between indirect and direct FDS and the strength of stochastic processes in natural populations coevolution, two main outcomes occur: trench warfare dynamics as the long-term maintenance of genetic and phenotypic diversity and an arms race with ultimate fixation of genotypes in both species. Coevolution is a prime example of eco–evo processes in which there is feedback between short-term ecological dynamics and the evolutionary dynamics of allele frequencies. With new sequencing technologies, it is possible to reveal how coevolution shapes the genomic diversity and genome evolution of antagonistic species.

eLS subject area: Evolution & Diversity of Life

How to cite:
Tellier, Aurélien; Brown, James KM; Boots, Mike; and John, Sona. *Theory of Host–Parasite Coevolution: From Ecology to Genomics*, eLS, **Vol 2**: 1–10, 2021.
DOI: 10.1002/9780470015902.a0029063

Introduction

In every natural ecosystem where multiple species share their environment, interspecific interactions are the norm rather than the exception. Indeed, most species interact with species from the same, lower and higher trophic levels, and in particular, multicellular organisms interact with a multitude of microbes constituting their microbiome. Interspecific interactions are generally classified into (1) mutualistic symbiosis, when reciprocal fitness benefits are cooperatively shared between the interacting species (e.g. plants and mycorrhizal fungi, and coral polyps and photosynthetic dinoflagellates), (2) antagonistic, when one species' fitness increases at the detriment of the other species' fitness (e.g. hosts and pathogens, parasites and parasitoids) or (3) commensal, when one organism benefits, while the other is generally unaffected.

Mutualistic and antagonistic interactions are in particular characterised by coevolution, that is the reciprocal adaptation of one species to the other over time. Interactions between species can also move along the antagonistic–commensal–mutualistic continuum depending on environmental conditions as well as the surrounding species community (Thompson, 2005). Interspecific coevolution is strict when the interaction is strong and exclusive in space and time, for example when the life cycle of one species depends strictly on another. Conversely, coevolution can be diffuse when the interaction is heterogeneous in space or varies in intensity over time, for example when a parasite can infect many host species. Strict antagonistic interaction exerts strong selective pressure on each partner. This is expected to result in rapid reciprocal adaptation facilitated by an increase in recombination and mutation rates. Coevolutionary processes are thus of paramount importance for shaping species diversity at all trophic levels (Schmid-Hempel, 2013). **See also: Interspecific Interaction**; **Reproductive Parasitism and Positive Fitness Effects of Heritable Microbes**; **Coevolution: Host-Parasite**; **Parasitism: The Variety of Parasites**; **Red Queen Hypothesis**

In this article, we focus on antagonistic coevolution in which one species (parasite or pathogen) decreases the fitness of its host, leading to reciprocal adaptation of infection strategies by the parasites counteracted by resistance mechanisms in the host.

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

In host–parasite coevolution, the traits under strong selection are the parasite *infectivity* and *virulence* and the innate and adaptive immune systems of the host (defining *resistance*). Host–parasite coevolution influences the likelihood of (1) outbreaks of diseases caused by new infectious and virulent parasite genotypes, (2) new parasites which generate pandemics of emerging diseases and zoonoses in the human population, (3) pandemics in farm animals and (4) new, sometimes damaging, strains of diseases in crops. As these are the major challenges facing human society, understanding the evolutionary mechanisms and finding the genetic basis of host–parasite coevolution is beneficial to human health and agriculture (Dieckmann *et al*., 2005; Schmid-Hempel, 2013; Milgroom, 2015). The innate immune system is close to the universal feature of multicellular organisms, and there are fascinating similarities between plants and animals (invertebrates as well as vertebrates) (Dodds and Rathjen, 2010; Jones *et al*., 2016). With advances in microbiome studies and metagenomics, it appears that the immune system in plants and animals (including humans) critically determines the host regulation of its microbiome composed of multiple commensals, mutual symbionts and also antagonistic microbes (Eberl and Pradeu, 2018). Long-term interactions between hosts and their parasites therefore play a major role in shaping the genome of all species. **See also: Red Queen Theory**; **Population Biology of Plant Pathogens**; **Evolution of Adaptive Immunity**; **Plant Innate Immunity**; ***Drosophila* Innate Immunity**; **Major Histocompatibility Complex (MHC) Genes: Evolution**; **Innate Immune Mechanisms: Nonself Recognition**

With advances in sequencing technology, it is now possible to study the genomic and genetic determinants of coevolution at unprecedented resolution, for example by sampling and sequencing several natural populations of hosts and parasites (Feurtey *et al*., 2016; Karasov *et al*., 2018; Ebert and Fields, 2020) and also by experimental coevolution in controlled conditions (Retel *et al*., 2019; Papkou *et al*., 2019). In the light of advances in genomics, this article (1) summarises the theoretical background of coevolutionary models and (2) links ecological dynamics to their effects on the genomes of hosts and parasites. There is a large and growing literature on coevolutionary models, and here, we explain the processes underlying coevolution common to all models. We describe the outcomes and dynamics of coevolution as the result of simple selective and stochastic processes and how they link to ecological features of the interaction. We conclude by providing an outlook on the effects of genomic signatures of coevolution which link ecology and genomics.

Overview of the Main Models of Coevolution

Defining the infection matrix and fitness costs

Models of coevolution make assumptions about the interaction between several genotypes of hosts and parasites within a population (Haldane and Jayakar, 1963; Leonard, 1977; Frank, 1992). Interactions between different pairs of host and parasite genotypes ($G \times G$ interactions) produce different phenotypic outcomes such as successful infection and host susceptibility or conversely failed infection and host resistance. The link between the $G \times G$ encounter and the phenotype is summarised by an infection matrix in which each row is a host genotype, and each column a parasite genotype (Dybdahl *et al*., 2014). The numbers in the cells of the matrices in **Table 1** indicate the phenotype of a $G \times G$ interaction ranging from 0 (no infection, total resistance) to 1 (fully successful infection, complete susceptibility). In a simple case with only two host and parasite genotypes (**Table 1**), the gene-for-gene (GFG) model of interaction is common in many plant–parasite systems, and the matching-allele (MA) model is present in some animal–parasite systems (**Table 1**; Dybdahl *et al*., 2014). The infection matrix thus defines the specificity of the host–parasite genotype interaction, namely the *infectivity* of the parasite and the *qualitative resistance* of the host. Simple models such as the GFG and MA are obtained under the assumption that only few major genotypes are present in the population. Conversely, the specificity traits, infectivity and resistance, can also be quantitative combinations of multiple minor genotypes and represented as a high-dimensional infection matrix (**Table 1**; Gandon and Michalakis, 2002; Boots *et al*., 2014).

In most models, four or five parameters are used to describe the host–parasite interaction. *Disease severity* is the amount of damage inflicted by the parasite upon an infected host; it is often measured as the proportion by which successful infection reduces host fitness, such as the production of offspring. In evolutionary (animal and human) epidemiology, *virulence* is the increased death rate of infected hosts. (Note that virulence is often used in plant pathology in a completely different sense, describing the specificity of interaction between a parasite and a susceptible host plant. We use the term *infectivity* for this type of specificity, as it is commonly used in the evolutionary literature (Sacristan and Garcia-Arenal, 2008).) Especially in plant pathology, there is variation in the *effectiveness of qualitative resistance* against *noninfective* parasites in GFG interactions; most GFG resistances used in plant breeding have high effectiveness. Finally, there is a *cost of resistance* and a *cost of infectivity* for the host and parasite, respectively. It is assumed that there are trade-offs in the host, such that resistant host genotypes have lower mean fitness than susceptible hosts in the absence of the parasite (Tian *et al*., 2003). Similarly, it is assumed that infective parasites with a wider range of susceptible hosts genotypes have lower mean fitness (Thrall and Burdon, 2003; Montarry *et al*., 2010). The latter two costs are generally small and thus difficult to detect experimentally. **See also: Fitness Costs of Plant Disease Resistance**; **Virus–Plant Co-evolution**; **Avirulence Genes**

Discrete time population genetics models

Mathematical modelling of coevolution has been the focus of much attention since the early days of evolutionary genetics, following the pioneering work by Haldane and Jayakar (1963). These models assume infinite (or very large) population sizes of hosts and parasites, discrete time (discrete generations without overlapping of parents and offspring) and a frequency-dependent disease transmission which depends only on the genotype

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

Table 1 Defining infection matrices for major genes (α_M) or minor genes (α_m) and examples

General definition of the infection matrix	Numerical examples of matrices
$\alpha_M = \begin{pmatrix} 1 & \alpha_{1,2} \\ \alpha_{2,1} & 1 \end{pmatrix}$	$\alpha_{M\_\mathrm{GFG}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ or $\alpha_{M\_\mathrm{MA}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\alpha_m = (\alpha_{i,j})$	$\alpha_{m\_\mathrm{ex}} = \begin{pmatrix} 1 & 0.1 & \dots \\ 0.2 & 0.9 & \dots \\ \vdots & \vdots & \dots \end{pmatrix}$

Note: Major genes can exhibit a simple 2×2 genotype interaction, and the major models include the gene-for-gene (α_{M_GFG}) or the matching allele (α_{M_MA}). For quantitative traits with many minor genes, the matrix (α_m) has higher dimension with $A_{\mathrm{host}} \times A_{\mathrm{parasite}}$ genotypes, and a numerical example is given (α_{m_ex}). If many genotypes are possible in each species, with a maximum of A_H host and A_P parasite genotypes, the matrix has dimension $A_\mathrm{H} \times A_\mathrm{P}$, and $\alpha_{i,j}$ (with $i \in \{1, \dots, A_\mathrm{H}\}$ and $j \in \{1, \dots, A_\mathrm{P}\}$) determines the infection success (phenotype) of a given pairwise interaction.
Dybdahl *et al*. (2014); Gandon and Michalakis (2002); and Boots *et al*. (2014).

frequency in each population. In a simple model (Model A), the mean fitness of a host genotype is the product of its own fitness cost, the frequency of parasites which are genetically capable of infecting it, the frequency with which those parasites infect it and the disease severity caused by infection. As usual in population genetics (Charlesworth and Charlesworth, 2010), a host genotype's fitness is then standardised by dividing by the mean fitness across all host genotypes, weighted by the genotype frequencies. Multiplying host genotype frequencies by their respective fitnesses gives the genotype frequencies in the next discrete generation. This scenario is often modelled by difference equations which describe changes from one generation to the next (e.g. one year to the next of an annual crop plant and a pathogen). This model can be extended to include more realistic features of hosts and parasites (Leonard, 1977; Tellier and Brown, 2007).

A simple discrete time model (Model A) describes the dynamics of A_H and A_P genotypes in the host and parasite over time

$$H_{i,g+1} = H_{i,g}(1 - c_{\mathrm{H}i}) \sum_{j=1}^{A_\mathrm{P}} P_{j,g}(1 - \alpha_{i,j} + (1-s)\alpha_{i,j})/W_\mathrm{H} \quad (1)$$

$$P_{j,g+1} = P_{j,g}(1 - c_{\mathrm{P}j}) \sum_{i=1}^{A_\mathrm{H}} H_i \alpha_{i,j}/W_\mathrm{P} \quad (2)$$

where $H_{i,g}$ is the frequency of the host genotype i in generation g, $P_{j,g}$ is the corresponding frequency of parasite genotype j, W_H and W_P are the host and parasite population fitnesses, respectively, at the end of generation g, s is the disease severity, $c_{\mathrm{H}i}$ is the cost of resistance of host genotype i, $c_{\mathrm{P}j}$ is the cost of infectivity of parasite genotype j and $\alpha_{i,j}$ is the infectivity of parasite genotype j on host genotype i. This is essentially the same as the model without direct frequency-dependent selection (FDS) in Tellier and Brown (2007) and is similar to the model of Leonard (1977).

Continuous time epidemiological models

Another type of model develops early work on infectious disease epidemiology in human populations pioneered by Bernoulli and D'Alembert in the eighteenth century, followed by Kermack and McKendrick (1927) and more recently by Anderson and May (1979). These models, known as SI (susceptible infected) or SIR (susceptible infected resistant (removed-recovered)), assume that the disease transmission is density dependent, that is it depends on the proportion of infected individuals in a population. The models compute the changes in number (or densities) of infected and noninfected hosts over time in a continuous time setting using differential equations (Model B). This type of model is especially well adapted to parasites which undergo many generations per host generation and have a life cycle that depends on the host for survival and reproduction. It is also suitable for following epidemics within a season in a crop field. **See also: Epidemiology of Plant Disease**

The following continuous time model (Model B) includes several host and parasite genotypes (May and Anderson, 1983; Gandon *et al*., 2002; Tellier and Brown, 2009; Boots *et al*., 2014):

$$\frac{\mathrm{d}H_i}{\mathrm{d}t} = H_i \left[b(1 - c_{\mathrm{H}i}) - d - \sum_{j=1}^{A_\mathrm{P}} \beta \alpha_{i,j}(1 - c_{\mathrm{P}j}) \sum_{k=1}^{A_\mathrm{H}} I_{k,j} \right] + b(1 - c_{\mathrm{H}i})(1-s) \sum_{j=1}^{A_\mathrm{P}} I_{i,j} \quad (3)$$

$$\frac{\mathrm{d}I_{i,j}}{\mathrm{d}t} = I_{i,j}(-d - \delta) + \beta H_i \alpha_{i,j}(1 - c_{\mathrm{P}j}) \sum_{k=1}^{A_\mathrm{H}} I_{k,j} \quad (4)$$

where H_i is the number of healthy host individuals of genotype i in the population, $I_{i,j}$ is the number of infected hosts of type i infected by parasite of genotype j, β is the disease transmission rate, δ is the pathogen virulence and the host has a birth rate b and a natural death rate d (in the absence of the parasite). Other parameters are as in Model A above.

Two Types of Frequency-Dependent Selection

Most theoretical models of coevolution generate coevolutionary dynamics which describe the change in genotype frequencies over time in the antagonistic species due to reciprocal adaptation. For models both in discrete time (A) and continuous time (B), there is a polymorphic equilibrium state at which all genotypes

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

are maintained in host and parasite populations. Both populations are then polymorphic for the coevolving traits. Alternatively, the populations of the host or parasite or both may become fixed for one genotype and thus be monomorphic. A large body of literature has focused on the conditions to reach and maintain a polymorphic equilibrium with several genotypes.

The coevolutionary dynamics result from the effect of two types of *FDS*. FDS occurs when the fitness of a genotype, that is its advantage or disadvantage compared to other genotypes, depends on the frequency of the same genotype or another genotype. The first selective force which occurs in all coevolution models is negative indirect frequency-dependent selection (niFDS) under which the fitness of a genotype depends on genotype frequencies in the antagonistic species. In a simple MA model, for example (**Table 1**) the fitness of host type 1 is high when the frequency of the corresponding parasite genotype 1 is low. As the frequency of host type 1 increases, the frequency of parasite type 1 also rises because a parasite genotype is fitter when corresponding, susceptible host genotypes are more common. When the frequency of host type 1 is high, it will have selected a high frequency of parasite type 1. This in turn reduces the fitness of host type 1, which then falls. Lastly, the fall in frequency of host type 1 reduces the fitness of parasite type 1, which falls once again to a low frequency (Leonard, 1977; Frank, 1992). In the simple, discrete time model A, the only selective forces are niFDS in the host and parasite (Tellier and Brown, 2007).

The second selective force, negative direct frequency--dependent selection (ndFDS), occurs when the fitness of a genotype is a negative function of its own frequency so that when a genotype becomes common, its fitness is significantly decreased. ndFDS is generated by several ecological factors and life-history traits (review in Brown and Tellier, 2011). Note that ndFDS is not generated by the costs of resistance or infectivity, because these costs do not depend on the genotype frequencies. The factors promoting ndFDS include (1) polycyclic diseases with high autoinfection rates (Tellier and Brown, 2007), (2) seed banking in hosts or perennial structures such as fungal mycelium in parasites (Tellier and Brown, 2009; Verin and Tellier, 2018), (3) feedbacks between ecology and evolution (eco–evo) which result from the continuous time epidemiological model B (Tellier and Brown, 2009; Ashby and Boots, 2017; Ashby *et al*., 2019, see the following discussion for details) and (4) heterogeneous selection in space, such as varying disease severity or disease prevalence or costs of resistance or infectivity (Gavrilets and Michalakis, 2008; Tellier and Brown, 2011).

The respective strengths of the two types of FDS determine the deterministic dynamics of allele frequencies and the amplitude and period of genotype frequency cycles in time. There are three possible outcomes: stable equilibrium (**Figure 1a**), unstable equilibrium (**Figure 1b**) and stable cycles (**Figure 1c**). The first scenario occurs when ndFDS is stronger than niFDS, and fluctuations (cycling) of genotype frequencies converge towards the polymorphic equilibria shown as dotted lines in **Figure 1a**. The equilibria are stable, and polymorphism of several genotypes is expected to be maintained over the long term in both host and parasite populations. The second scenario occurs when niFDS is stronger than ndFDS, generating cycles of allele frequencies which increase in amplitude over time, until they reach the maximum (1) or minimum (0) frequency in the population, and one of the genotypes is thus lost (**Figure 1a**). In this case, the polymorphic equilibria for hosts and parasites, shown as dotted lines in **Figure 1a**, are not reached, and the frequencies diverge away from these values over time. The equilibrium is thus unstable. Finally, in a third case, when neither niFDS nor ndFDS clearly dominates, and with specific, narrow parameter values, endless 'stable' cycles with fixed amplitude and period can be observed. Polymorphism is maintained in a dynamic way over time.

The polymorphic equilibrium point is defined by the parameters of the model (Frank, 1992; Brown and Tellier, 2011), and in simple models, it is possible to determine analytically the stability of the internal equilibrium point from the model parameters. Examples of such computations are found for discrete time (e.g. Model A) in Leonard (1977), Tellier and Brown (2007), and for continuous time (e.g. Model B) in May and Anderson (1983); Živković *et al*. (2019). In summary, if niFDS is the key force generating coevolutionary dynamics, the costs of resistance and infectivity are necessary but not sufficient to promote a stable equilibrium. Stable polymorphism can only appear if ndFDS occurs and is strong enough.

Importance of Stochasticity

Deterministic models of an interacting host and parasite predict general conditions for stability of the polymorphic equilibria in terms of the type of FDS, but applying these results to real datasets is challenging. A successful method has been to study changes in phenotypic traits over time and assessing their fit to stable or unstable coevolutionary dynamics (Gandon *et al*., 2008). However, in trying to decipher the genetic and genomic signatures of coevolutionary dynamics, it is necessary to account for stochastic processes common in nature. There are three main sources of stochasticity: random genetic drift, stochastic disease transmission between individuals and environmental variation in space and time.

Genetic drift is due to an element of chance in reproduction resulting in a variable number of offspring for each individual (Charlesworth and Charlesworth, 2010). Genetic drift is not expected to change mean allele frequency trajectories, but it nudges frequencies off the expected value at the next generation. This produces cyclical fluctuations of allele frequencies around the deterministic expectations (Tellier and Brown, 2007). Genetic drift can lead to fixation or loss of alleles and thus reduces the stability of polymorphism in host and parasite populations (Kirby and Burdon, 1997, Tellier *et al*., 2014). The effect of genetic drift is stronger, that is there is faster fixation of alleles, when the effective population size is small. Thus, under demographic scenarios such as population bottlenecks, genetic drift can result in random fixation or extinction of alleles and loss of polymorphism (Gokhale *et al*., 2013).

Another source of stochasticity in a population is the disease transmission process itself (Salathé *et al*., 2005; Allen, 2008, 2017). The transmission rate in a deterministic model is

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

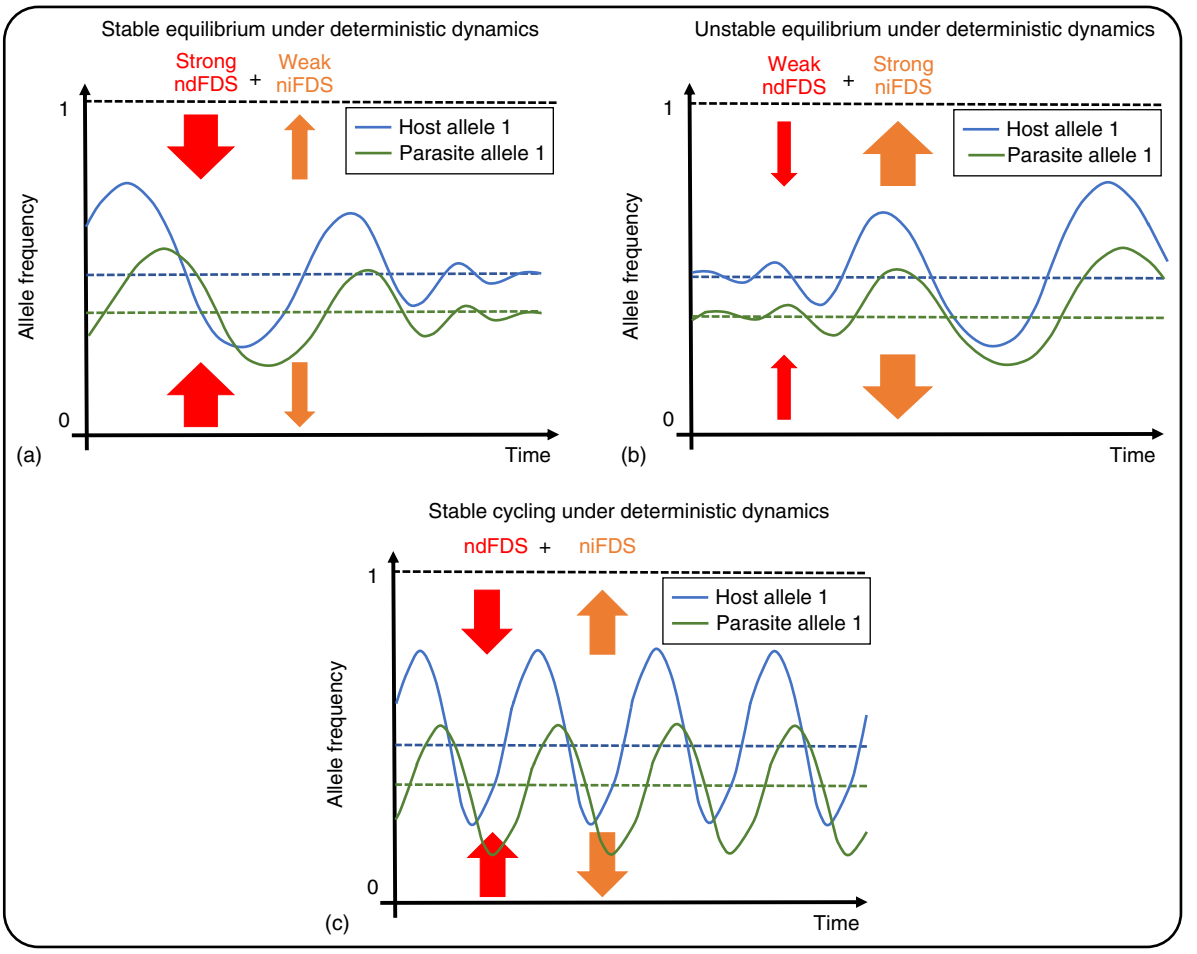

Figure 1 Schematic representation of the possible coevolutionary dynamics under deterministic models. (a) The polymorphic equilibrium point is stable so that allele frequencies converge towards this value (dotted line for host in blue, parasite in green). A stable equilibrium is possible when ndFDS is strong compared to niFDS. (b) The polymorphic equilibrium point is unstable so that allele frequencies diverge away from this value (dotted line for host in blue, parasite in green). An unstable equilibrium is obtained when niFDS is strong compared to ndFDS. (c) The special case where the polymorphic equilibrium point has stable cycles of genotype frequencies with given period and amplitude. In a deterministic model, this can only happen with specific, narrow sets of parameter values.

an average value, but in fact, an infected individual can randomly infect more or fewer individuals. A 'superspreader' can infect many individuals. Disease transmission can vary spatially depending on the level of connectivity between populations, countries and continents, resulting in the introduction of pathogen to a completely susceptible population or in superspreaders with many contacts. In human populations, age structuring and grouping may create heterogeneity in transmission as people of the same age group are more likely to be in contact (e.g. in schools and in old people's homes). The health status of different individuals also affects the transmission probability as disease severity can vary greatly between individuals. These factors need to be considered in a realistic model to predict transmission dynamics. Stochasticity in disease transmission can be modelled using a contact tracing model with a branching process. This assumes that at the start of the epidemic most individuals in a population are susceptible, but as it progresses, contacts with individuals who are already infected will increasingly not produce new infections (Allen, 2008, 2017). Such effects have not been quantified in coevolutionary models, but analytical solutions exist for simple SI and SIR models (Allen, 2017).

Finally, especially in diseases of plants and invertebrates, the abiotic environment (temperature, humidity, light regime, availability of food for the host, etc.) plays a major role in determining infection success, disease severity and parasite virulence. The periodic or seasonal pattern of many infectious diseases depends strongly on environmental factors such as weather conditions. Variable, complex Host × Parasite × Environment interactions (Laine and Tellier, 2008; Parratt *et al.*, 2016) can vary the infection matrix in time and also in space across populations (see the Geographic Mosaic of coevolution, Thompson, 2005). Overall, the effect of stochastic processes is to drive genotype frequencies towards fixation or loss, unless ndFDS is strong compared to stochastic effects (**Figure 2**).

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

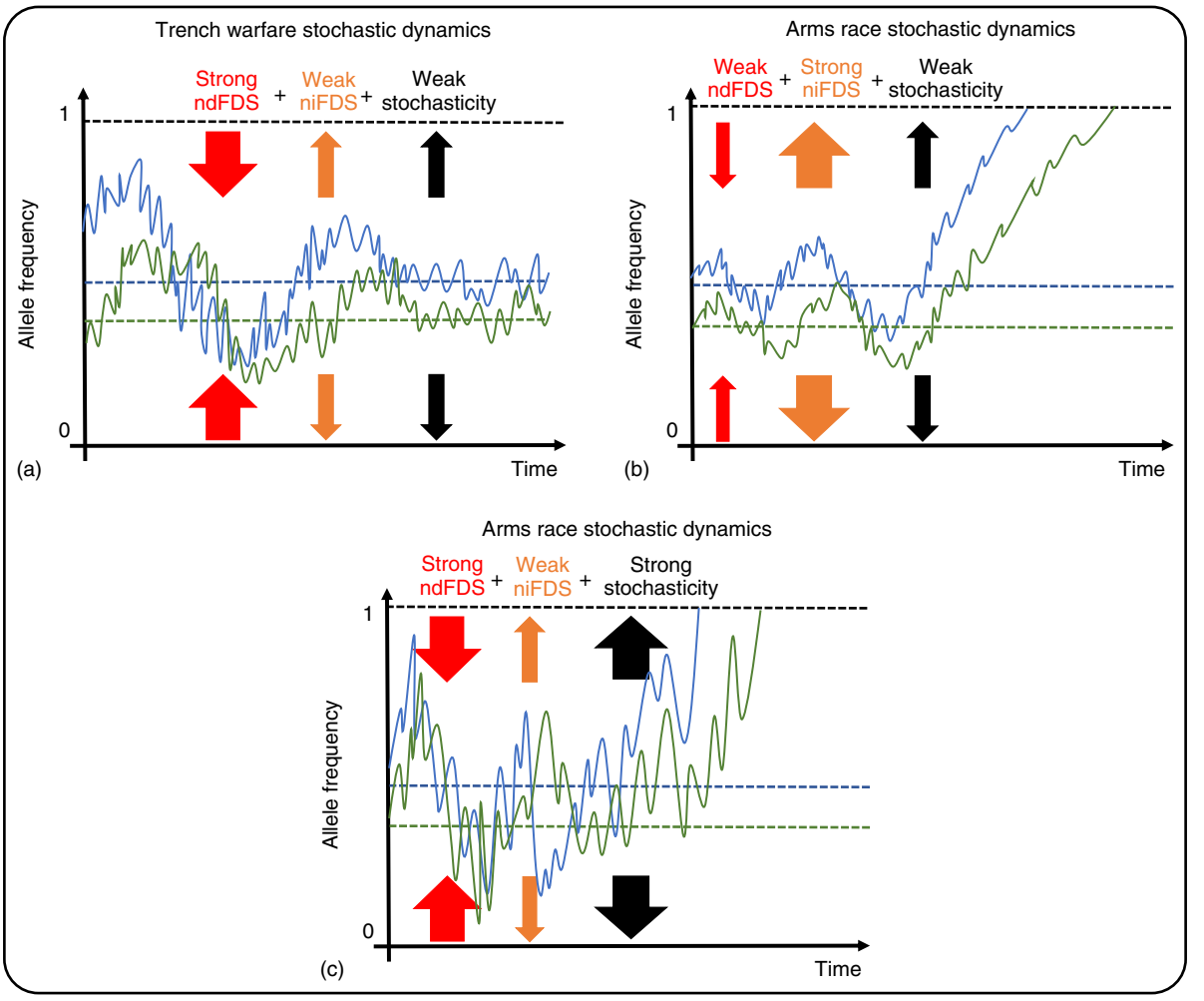

Figure 2 Schematic representation of the possible coevolutionary dynamics with stochastic processes occurring in host and parasite populations. (a) Allele frequencies nudge around the stable polymorphic equilibrium value from **Figure 1a**. Trench warfare dynamics occur when the stochastic processes are weak compared to ndFDS, so that as stochasticity nudges frequencies away from the equilibrium point, ndFDS forces them back towards the stable equilibrium value.Stahl *et al.* (1999); Holub (2001); and Woolhouse *et al.* (2002). (b) Allele frequencies diverge rapidly from the unstable polymorphic equilibrium point from **Figure 1b**, resulting in the rapid fixation of one host and one parasite genotype (arms race dynamics). (c) Allele frequencies vary wildly around the stable polymorphic equilibrium point from **Figure 1a** or **c** because stochastic processes are stronger than ndFDS, resulting in the fixation of one host and one parasite genotype (arms race dynamics). Modified from Tellier *et al.* (2014).

Thus, deterministic models are a valid approximation for an infinitely large, well-mixed and homogeneous population, but real finite-size populations can produce new phenomena because of the intrinsic stochasticity of the system. If stochastic processes are weak, then if a stable equilibrium occurs in the deterministic model, long-term polymorphism can be maintained in a finite population with some variation in allele frequencies (quasi-stable polymorphism) (**Figure 1a**). This situation is described as 'trench warfare' coevolutionary dynamics (**Figure 2a**, Stahl *et al.*, 1999; Holub, 2001; Woolhouse *et al.*, 2002). When stochastic processes are strong, however, for example when effective population sizes of hosts or parasites are small, then even when a deterministic model predicts a stable equilibrium (or stable cycles) (**Figure 1a,c**), stochastic fluctuations can push the system away from this stable point and cause fixation of genotypes, generating an 'arms race' scenario (**Figure 2c**; Tellier *et al.*, 2014). Alternatively, when an unstable equilibrium is predicted by a deterministic model (**Figure 1b**), stochasticity causes more rapid fixation of genotypes (**Figure 2b**). The introduction of new genotypes by mutation balances the effect of genetic drift and may trigger new coevolutionary cycles especially when alleles are already fixed (**Figure 2b,c**). An arms race can result from recurrent fixation of host and parasite alleles over time (Bergelson *et al.*, 2001; Holub, 2001; Woolhouse *et al.*, 2002), and the time between fixation of alleles and new mutations depends on the population mutation rate, that is the effective population size multiplied by the mutation rate (Tellier *et al.*, 2014). **See also:** **Drift: Introduction**; **Population Genetics of Plant Pathogens**; **Stationary Allele Frequency Distributions**

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

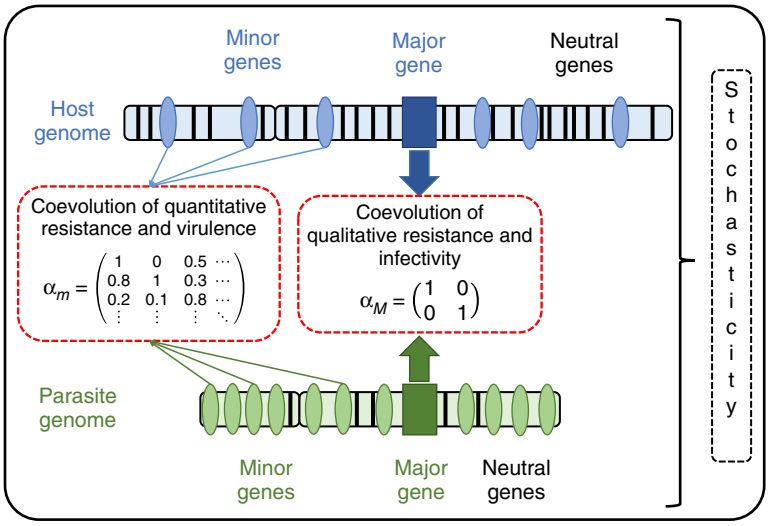

Figure 3 Schematic representation of the genomes of a host (in blue, top) and its parasite species (in green, bottom). The genome is composed of (1) major genes involved in coevolution (large squares), (2) many minor genes under coevolution (ellipsoid shapes) and (3) neutral or nearly neutral genes with respect to coevolution (black lines). Major genes may be under strict and strong coevolutionary dynamics and code for qualitative resistance in hosts and infectivity in parasites. The outcome of infection between different genotypes at those major loci is summarised as the small dimension infection matrix (α_M). Furthermore, host individuals vary in their quantitative resistance encoded by many minor genes. The parasite population exhibits variability for virulence and disease severity as quantitative traits determined by many minor genes. The quantitative interaction is summarised by a high-dimensional matrix (α_m) of the multilocus genotypes in the host and parasites. Finally, many genes in the host genome behave in a neutral manner with respect to coevolution. In the parasite genome, we expect fewer genes to be neutral (or nearly neutral) with respect to coevolution. All genes are subject to the effect of stochastic processes (genetic drift, demographic history, stochastic disease transmission and coevolution strength varying in space and time randomly) occurring in host and parasite populations.

Eco–Evo Feedbacks Due to Coevolution and Their Consequences

Finally, we consider some consequences of eco–evo feedbacks typical of coevolutionary dynamics as described by continuous time models such as Model B, characterised by the interplay between ecological and evolutionary dynamics. The ecological dynamics involve both changes in the population sizes of host and parasite which are impacted by the infection process. This results in allele frequency changes that define the evolutionary dynamics. In turn, changes in fitness affect population size, positively or negatively. By contrast, the simple, discrete time Model A only reflects the evolutionary dynamics of allele frequencies. Typically, eco–evo feedbacks generate strong ndFDS because the fitness of a host allele often negatively decreases with its own frequency due to its impact on the ecology that then feeds back into its fitness: the eco–evo feedback (Tellier and Brown, 2009; Ashby and Boots, 2017; Ashby *et al*., 2019). A first interesting outcome of complex eco–evo feedbacks is that they generate a large variety of coevolutionary dynamics (Ashby *et al*., 2019), depending on the life cycles of hosts and parasites. Second, eco–evo feedbacks generate varying population sizes of host and parasite in time, which affects the genome-wide, neutral and nonneutral, variation including single nucleotide polymorphisms (SNPs) (Živković *et al*., 2019).

Third, eco–evo feedbacks determine the genetic architecture of qualitative resistance and infectivity and quantitative resistance and virulence traits. Boots *et al*. (2014) developed a model analogous to Model B with many minor loci (genotypes) under different assumptions regarding the values of the infection matrix coefficients and trade-offs (costs of resistance and infectivity). The genotypic architecture of the host–parasite interaction, that is how many genes are involved and how strict is the interaction, evolves through eco–evo feedbacks and trade-offs driving the coevolutionary dynamics. The architecture of host resistance and parasite infectivity emerges as monomorphic (one optimal value), dimorphic (two values coexist) or polymorphic (several values can coexist), depending on the shape of trade-offs and the nature of the infection matrix. When linking these results to genetics, one can speculate that monomorphic or dimorphic traits are coded by major genes, while polymorphic traits are coded by minor loci (Boots *et al*., 2014). As such, this theoretical framework allows formulation of links between ecological conditions, host and parasite trade-offs and the genetic architecture of resistance and infectivity (number of major and minor genes involved), although these predictions have yet to be tested at the genomic and phenotypic levels.

The Age of Genomics

So far, we have only used the generic term ‘genotypes’ as the genetic determinants of the coevolutionary dynamics, and we now define more precisely the genetic and genomic basis of

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

host–parasite interactions. With regard to host–parasite coevolution, we can partition the genomes of the antagonistic species as follows (**Figure 3**). First, host genomes contain only a few classes of major gene subject to coevolution, but these loci can be under strong coevolution when they interact with allelic variants at the parasite's major infectivity genes. The result of the G × G interaction between alleles defines the infectivity/resistance phenotype as described in the GFG or MA models (**Table 1**). At the molecular level, an important class of such genes in invertebrates and especially in plants encodes nucleotide-binding site and leucine-rich repeat containing receptors (NLRs) which detect the presence of pathogens and thence trigger defence responses (Dodds and Rathjen, 2010; Jones *et al*., 2016). In plants, NLR genes are widely used in breeding crop varieties, although the resistance is usually temporary because pathogens usually evolve infectivity specific to a particular GFG resistance more or less rapidly (Dodds and Rathjen, 2010). In vertebrates, the major histocompatibility complex, which is involved in initial recognition of pathogens and the subsequent activation of adaptive immunity, is a prime candidate for major determinants of coevolution, as are interferon genes involved in virus resistance (Ansari *et al*., 2017). Other major resistance genes encode host ribonucleic acid (RNA) providing defence against viruses (Obbard *et al*., 2009) or fungal pathogens (Dunker *et al*., 2020). The mutation mechanisms in these major loci can be point mutations changing amino acids, and also inactivation of susceptibility factors by insertion–deletion or transposable element activity, or changes in gene expression. Second, host genomes typically contain many quantitative loci of resistance scattered over the genome, each having a small quantitative effect. These loci thus generate a large number of genotypes differing from one another by the amount of resistance. In plants, some of these minor genes may interact in a G × G manner with parasite genes (**Table 1**) but most apparently do not, so the resistance they encode is usually durable (Cowger and Brown, 2019). At the molecular level, the genes responsible are poorly understood but may largely control less-specific quantitative defence mechanisms (Poland *et al*., 2009). Third, a large portion of the host genome consists of loci with neutral (or almost neutral) genes, involved in other life-history and physiological traits, which have little effect on interactions with parasites.

Many parasite genomes contain a large number of genes which determine interactions with the host. Many but not all of these genes in bacterial and fungal pathogens are effectors which manipulate the host's physiology or target the immune system (Rovenich *et al*., 2014; Toruño *et al*., 2016). There are many examples of such genes, known as avirulence genes (in plant pathology), in crop pathogens which determine a GFG interaction (Rovenich *et al*., 2014). Parasite genomes also contain many quantitative loci determining the level of virulence and disease severity, for example defining the amount of resources taken from the host and thus the rates of parasite reproduction and disease transmission. The evolution of virulence has been investigated in great detail (Dieckmann *et al*., 2005) and is influenced by the polymorphism and eco–evo dynamics of major genes (effectors) (Gandon *et al*., 2002). As many parasites are strongly dependent on their hosts for survival, growth and reproduction, it is expected that few parasite genes evolve in a near-neutral manner.

Conclusion

Antagonistic coevolution is a multifaceted and complex process which affects the whole genome of hosts and parasites, including those of medical and agricultural importance. A great deal of effort has therefore been spent on deciphering the molecular basis of coevolution in hosts and parasites. We have presented an overview of key points of the theory of coevolution and its interaction with random processes, including genetic drift and disease transmission. The resulting coevolutionary dynamics can be observed in the genome at different loci, and new methods using joint analysis of host and parasite genome-wide data will allow detailed tests of the models.

Glossary

Eco–evo dynamics The interaction between ecological and evolutionary processes and time scales which can generate rapid evolutionary responses to ecological changes.

Frequency-dependent selection (FDS) A form of natural selection in which the fitness of a genotype depends on its own current frequency in the population or the frequency of another gene in the same or another species.

Genetic drift Stochastic changes in allele frequency due to randomness in the birth, reproduction and death process of individuals.

Infectivity Parasite or pathogen's ability to successfully infect a given host. This is often called virulence in plant pathology.

Resistance Host's ability to partially or completely stop infection.

Virulence In animals, including humans, the severity of disease or the disease-induced death rate of the host.

References

Allen LJ (2008) An introduction to stochastic epidemic models. In: *Mathematical Epidemiology*, pp 81–130. Berlin, Heidelberg: Springer.

Allen LJ (2017) A primer on stochastic epidemic models: formulation, numerical simulation, and analysis. *Infectious Disease Modelling* **2** (2): 128–142.

Anderson RM and May RM (1979) Population biology of infectious diseases: part I. *Nature* **280** (5721): 361–367.

Ansari MA, Pedergnana V, Ip CLC, *et al*. (2017) Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nature Genetics* **49** (5): 666–673.

Ashby B and Boots M (2017) Multi-mode fluctuating selection in host–parasite coevolution. *Ecology Letters* **20** (3): 357–365.

Ashby B, Iritani R, Best A, White A and Boots M (2019) Understanding the role of eco-evolutionary feedbacks in host–parasite coevolution. *Journal of Theoretical Biology* **464**: 115–125.

Bergelson J, Kreitman M, Stahl EA and Tian D (2001) Evolutionary dynamics of plant R-genes. *Science* **292** (5525): 2281–2285.

Boots M, White A, Best A and Bowers R (2014) How specificity and epidemiology drive the coevolution of static trait diversity in hosts and parasites. *Evolution* **68** (6): 1594–1606.

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

Brown JKM and Tellier A (2011) Plant-parasite coevolution: bridging the gap between genetics and ecology. *Annual Review of Phytopathology* **49**: 345–367.

Charlesworth B and Charlesworth D (2010) *Elements of Evolutionary Genetics*. Roberts and Co Publishers: Greenwood Village, CO, USA.

Cowger C and Brown JKM (2019) Durability of quantitative resistance in crops: greater than we know? *Annual Review of Phytopathology* **57**: 253–277.

Dieckmann U, Metz JA and Sabelis MW (eds) (2005) *Adaptive Dynamics of Infectious Diseases: In Pursuit of Virulence Management*, vol. **2**. Cambridge University Press.

Dodds PN and Rathjen JP (2010) Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* **11** (8): 539–548.

Dunker F, Trutzenberg A, Rothenpieler JS, *et al*. (2020) Oomycete small RNAs bind to the plant RNA-induced silencing complex for virulence. *eLife* **9**: e56096.

Dybdahl MF, Jenkins CE and Nuismer SL (2014) Identifying the molecular basis of host-parasite coevolution: merging models and mechanisms. *The American Naturalist* **184** (1): 1–13.

Eberl G and Pradeu T (2018) Towards a general theory of immunity? *Trends in Immunology* **39** (4): 261–263.

Ebert D and Fields PD (2020) Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics* **21** (12): 754–768.

Feurtey A, Gladieux P, Hood ME, *et al*. (2016) Strong phylogeographic co-structure between the anther-smut fungus and its white campion host. *New Phytologist* **212** (3): 668–679.

Frank SA (1992) Models of plant-pathogen coevolution. *Trends in Genetics* **8** (6): 213–219.

Gandon S and Michalakis Y (2002) Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *Journal of Evolutionary Biology* **15** (3): 451–462.

Gandon S, van Baalen M and Jansen VA (2002) The evolution of parasitevirulence, superinfection, and host resistance. *The American Naturalist* **159** (6): 658–669.

Gandon S, Buckling A, Decaestecker E and Day T (2008) Host–parasite coevolution and patterns of adaptation across time and space. *Journal of Evolutionary Biology* **21** (6): 1861–1866.

Gavrilets S and Michalakis Y (2008) Effects of environmental heterogeneity on victim–exploiter coevolution. *Evolution* **62** (12): 3100–3116.

Gokhale CS, Papkou A, Traulsen A and Schulenburg H (2013) Lotka–Volterra dynamics kills the Red Queen: population size fluctuations and associated stochasticity dramatically change host-parasite coevolution. *BMC Evolutionary Biology* **13** (1): 254.

Haldane JBS and Jayakar SD (1963) Polymorphism due to selection of varying direction. *Journal of Genetics* **58** (2): 237–242.

Holub EB (2001) The arms race is ancient history in *Arabidopsis*, the wildflower. *Nature Reviews Genetics* **2** (7): 516–527.

Jones JD, Vance RE and Dangl JL (2016) Intracellular innate immune surveillance devices in plants and animals. *Science* **354** (6316): aaf6395.

Karasov TL, Almario J, Friedemann C, *et al*. (2018) *Arabidopsis thaliana* and *Pseudomonas* pathogens exhibit stable associations over evolutionary timescales. *Cell Host & Microbe* **24** (1): 168–179.

Kermack WO and McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A* **115** (772): 700–721.

Kirby GC and Burdon JJ (1997) Effects of mutation and random drift on Leonard's gene-for-gene coevolution model. *Phytopathology* **87** (5): 488–493.

Laine AL and Tellier A (2008) Heterogeneous selection promotes maintenance of polymorphism in host–parasite interactions. *Oikos* **117** (9): 1281–1288.

Leonard KJ (1977) Selection pressures and plant pathogens. *Annals of the New York Academy of Sciences* **287** (1): 207–222.

May RM and Anderson RM (1983) Epidemiology and genetics in the coevolution of parasites and hosts. *Proceedings of the Royal Society of London. Series B. Biological Sciences* **219** (1216): 281–313.

Milgroom MG (2015) *Population Biology of Plant Pathogens*. St. Paul, MN: American Phytopathological Society Press.

Montarry J, Hamelin FM, Glais I, Corbière R and Andrivon D (2010) Fitness costs associated with unnecessary virulence factors and life history traits: evolutionary insights from the potato late blight pathogen *Phytophthora infestans*. *BMC Evolutionary Biology* **10** (1): 1–9.

Obbard DJ, Gordon KH, Buck AH and Jiggins FM (2009) The evolution of RNAi as a defence against viruses and transposable elements. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364** (1513): 99–115.

Papkou A, Guzella T, Yang W, *et al*. (2019) The genomic basis of Red Queen dynamics during rapid reciprocal host–pathogen coevolution. *Proceedings of the National Academy of Sciences* **116** (3): 923–928.

Parratt SR, Numminen E and Laine AL (2016) Infectious disease dynamics in heterogeneous landscapes. *Annual Review of Ecology, Evolution, and Systematics* **47**: 283–306.

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC and Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. *Trends in Plant Science* **14** (1): 21–29.

Retel C, Kowallik V, Huang W, *et al*. (2019) The feedback between selection and demography shapes genomic diversity during coevolution. *Science Advances* **5** (10): eaax0530.

Rovenich H, Boshoven JC and Thomma BP (2014) Filamentous pathogen effector functions: of pathogens, hosts and microbiomes. *Current Opinion in Plant Biology* **20**: 96–103.

Sacristan S and Garcia-Arenal F (2008) The evolution of virulence and pathogenicity in plant pathogen populations. *Molecular Plant Pathology* **9** (3): 369–384.

Salathé M, Scherer A and Bonhoeffer S (2005) Neutral drift and polymorphism in gene-for-gene systems. *Ecology Letters* **8** (9): 925–932.

Schmid-Hempel P (2013) *Evolutionary Parasitology: The Integrated Study of Infections, Immunology, Ecology, and Genetics*. Oxford, UK: Oxford University Press.

Stahl EA, Dwyer G, Mauricio R, Kreitman M and Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature* **400** (6745): 667–671.

Tellier A and Brown JKM (2007) Stability of genetic polymorphism in host–parasite interactions. *Proceedings of the Royal Society B: Biological Sciences* **274** (1611): 809–817.

Tellier A and Brown JK (2009) The influence of perenniality and seed banks on polymorphism in plant-parasite interactions. *The American Naturalist* **174** (6): 769–779.

Tellier A and Brown JKM (2011) Spatial heterogeneity, frequency-dependent selection and polymorphism in host–parasite interactions. *BMC Evolutionary Biology* **11** (1): 319.

eLS © 2021, John Wiley & Sons, Ltd. www.els.net

Tellier A, Moreno-Gámez S and Stephan W (2014) Speed of adaptation and genomic footprints of host–parasite coevolution under arms race and trench warfare dynamics. *Evolution* **68** (8): 2211–2224.

Tian D, Traw MB, Chen JQ, Kreitman M and Bergelson J (2003) Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423** (6935): 74–77.

Thrall PH and Burdon JJ (2003) Evolution of virulence in a plant host-pathogen metapopulation. *Science* **299** (5613): 1735–1737.

Thompson JN (2005) *The Geographic Mosaic of Coevolution*. Chicago, IL: University of Chicago Press.

Toruño TY, Stergiopoulos I and Coaker G (2016) Plant-pathogen effectors: cellular probes interfering with plant defenses in spatial and temporal manners. *Annual Review of Phytopathology* **54**: 419–441.

Verin M and Tellier A (2018) Host–parasite coevolution can promote the evolution of seed banking as a bet-hedging strategy. *Evolution* **72** (7): 1362–1372.

Woolhouse ME, Webster JP, Domingo E, Charlesworth B and Levin BR (2002) Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32** (4): 569–577.

Živković D, John S, Verin M, Stephan W and Tellier A (2019) Neutral genomic signatures of host–parasite coevolution. *BMC Evolutionary Biology* **19** (1): 1–11.

Further Reading

Borghans JA, Beltman JB and De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* **55** (11): 732–739.

Rabajante JF, Tubay JM, Ito H, *et al*. (2016) Host–parasite Red Queen dynamics with phase-locked rare genotypes. *Science Advances* **2** (3): e1501548.

Radwan J, Babik W, Kaufman J, Lenz TL and Winternitz J (2020) Advances in the evolutionary understanding of MHC polymorphism. *Trends in Genetics* **36** (4): 298–311.

Song Y, Gokhale CS, Papkou A, Schulenburg H and Traulsen A (2015) Host–parasite coevolution in populations of constant and variable size. *BMC Evolutionary Biology* **15** (1): 212.

Thrall PH and Burdon JJ (2002) Evolution of gene-for-gene systems in metapopulations: the effect of spatial scale of host and pathogen dispersal. *Plant Pathology* **51** (2): 169–184.

eLS © 2021, John Wiley & Sons, Ltd. www.els.net